

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
28 October 2004 (28.10.2004)

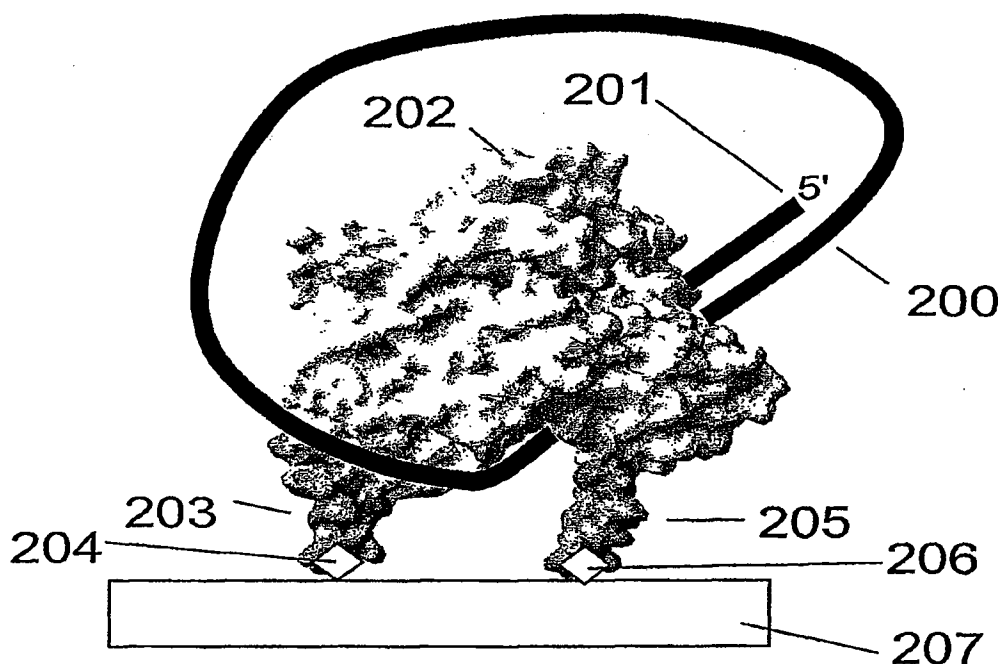
PCT

(10) International Publication Number
WO 2004/092331 A2

- (51) International Patent Classification⁷: C12N
- (21) International Application Number: PCT/US2004/010726
- (22) International Filing Date: 8 April 2004 (08.04.2004)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/461,522 8 April 2003 (08.04.2003) US
60/462,988 14 April 2003 (14.04.2003) US
- (71) Applicant (for all designated States except US): LI-COR, INC. [US/US]; Biotechnology Division, 4308 Progressive Avenue, Lincoln, Nebraska 68504 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): WILLIAMS, John, G., K. [US/US]; 6901 Pine Lake Court, Lincoln, Nebraska 68516 (US).
- (74) Agents: SNYDER, Joseph, R. et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, Eighth Floor, San Francisco, CA 94111-3834 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: COMPOSITION AND METHOD FOR NUCLEIC ACID SEQUENCING



(57) Abstract: The present invention provides compositions and methods for detecting incorporation of a labeled nucleotide triphosphate onto the growing end of a primer nucleic acid molecule. The method is used, for example, to genotype and sequence a nucleic acid. In a preferred embodiment, the method described herein detects individual NTP molecules.



Published:

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

COMPOSITION AND METHOD FOR NUCLEIC ACID SEQUENCING

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] The present invention claims priority to U.S. Patent Nos. 60/461,522 and
5 60/462,988, filed on April 8, 2003 and April 14, 2003, respectively, both of which are hereby
incorporated in their entirety for all purposes.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

10 [0002] The research embodied within the present application was funded in-part by the
Federal Government in research grant numbers R44 HG02292 and R44 HG02066. The
government may have certain rights in this application.

BACKGROUND OF THE INVENTION

15 [0003] The primary sequences of nucleic acids are crucial for understanding the function
and control of genes and for applying many of the basic techniques of molecular biology. In
fact, rapid DNA sequencing has taken on a more central role after the goal to elucidate the
entire human genome has been achieved. DNA sequencing is an important tool in genomic
analysis as well as other applications, such as genetic identification, forensic analysis, genetic
20 counseling, medical diagnostics, and the like. With respect to the area of medical diagnostic
sequencing, disorders, susceptibilities to disorders, and prognoses of disease conditions can
be correlated with the presence of particular DNA sequences, or the degree of variation (or
mutation) in DNA sequences, at one or more genetic loci. Examples of such phenomena
include human leukocyte antigen (HLA) typing, cystic fibrosis, tumor progression and
25 heterogeneity, p53 proto-oncogene mutations and ras proto-oncogene mutations (*see*,
Gyllenstein *et al.*, PCR Methods and Applications, 1: 91-98 (1991); U.S. Patent No.
5,578,443, issued to Santamaria *et al.*; and U.S. Patent No. 5,776,677, issued to Tsui *et al.*).

[0004] Various approaches to DNA sequencing exist. The dideoxy chain termination
method serves as the basis for all currently available automated DNA sequencing machines.
30 (*see*, Sanger *et al.*, Proc. Natl. Acad. Sci., 74: 5463-5467 (1977); Church *et al.*, Science, 240:
185-188 (1988); and Hunkapiller *et al.*, Science, 254: 59-67 (1991)). Other methods include

the chemical degradation method, (*see*, Maxam *et al.*, Proc. Natl. Acad. Sci., 74: 560-564 (1977), whole-genome approaches (*see*, Fleischmann *et al.*, Science, 269, 496 (1995)), expressed sequence tag sequencing (*see*, Velculescu *et al.*, Science, 270, (1995)), array-methods based on sequencing by hybridization (*see*, Koster *et al.*, Nature Biotechnology, 14, 1123 (1996)), and single molecule sequencing (SMS) (*see*, Jett *et al.*, J. Biomol. Struct. Dyn. 7, 301 (1989) and Schecker *et al.*, Proc. SPIE-Int. Soc. Opt. Eng. 2386, 4 (1995)).

[0005] U.S. Patent No. 6,255,083, issued to Williams and incorporated herein by reference, discloses a single molecule sequencing method on a solid support. The solid support is optionally housed in a flow chamber having an inlet and outlet to allow for renewal of reactants that flow past the immobilized polymerases. The flow chamber can be made of plastic or glass and should either be open or transparent in the plane viewed by the microscope or optical reader.

[0006] U.S. Patent No. 4,979,824, illustrates that single molecule detection can be achieved using flow cytometry wherein flowing samples are passed through a focused laser with a spatial filter used to define a small volume. Moreover, U.S. Patent No. 4,793,705 describes a detection system for identifying individual molecules in a flow train of the particles in a flow cell. The patent further describes methods of arranging a plurality of lasers, filters and detectors for detecting different fluorescent nucleic acid base-specific labels.

[0007] Single molecule detection on a solid support is described in Ishikawa, *et al. Jan. J. Apple. Phys.* 33:1571-1576. (1994). As described therein, single-molecule detection is accomplished by a laser-induced fluorescence technique with a position-sensitive photon-counting apparatus involving a photon-counting camera system attached to a fluorescence microscope. Laser-induced fluorescence detection of a single molecule in a capillary for detecting single molecules in a quartz capillary tube has also been described. The selection of lasers is dependent on the label and the quality of light required. Diode, helium neon, argon ion, argon-krypton mixed ion, and Nd:YAG lasers are useful in this invention (*see*, Lee *et al.* (1994) *Anal. Chem.*, 66:4142-4149).

[0008] The predominant method used today to sequence DNA is the Sanger method (*Proc. Natl. Acad. Sci.* 1977, 74, 5463) which involves use of dideoxynucleoside triphosphates as DNA chain terminators. Most high throughput-sequencing systems use this approach in combination with use of fluorescent dyes. The dyes may be attached to the terminator or be a part of the primer. The former approach is preferred as only the terminated fragments are

labeled. Multiplexing energy transfer fluorescent dyes are preferable over the use of single dyes.

[0009] U.S. Patent No. 6,306,607 describes modified nucleotides wherein the nucleotide has a terminally labeled phosphate, which characteristic is useful for single-molecule DNA sequencing in a microchannel. Using 4 different NTPs each labeled with a unique dye, real-time DNA sequencing is possible by detecting the released pyrophosphate having different labels. The cleaved PPi-Dye molecules are detected in isolation without interference from unincorporated NTPs and without illuminating the polymerase-DNA complex.

[0010] Despite the advances in U.S. Patent No. 6,255,083, a need currently exists for more effective and efficient compositions, methods, and systems for nucleic acid sequencing. Specifically, a need exists for improved nucleic acid sequencing compositions and methods to increase processivity. These and further needs are provided by the present invention.

SUMMARY OF THE INVENTION

[0011] The current invention provides compositions and methods to sequence nucleic acid. The compositions and methods allow for increasing the processivity index of polymerases and thus, results in more efficient nucleic acid sequencing. As such, in one aspect, the present invention provides a polymerase-nucleic acid complex, the polymerase-nucleic acid complex comprising: a target nucleic acid and a nucleic acid polymerase, wherein the polymerase has an attachment complex comprising at least one anchor which irreversibly associates the target nucleic acid with the polymerase for increasing the processivity index.

[0012] In one embodiment, the polymerase-nucleic acid complex further comprises a primer nucleic acid which complements a region of the target nucleic acid. In another embodiment, the attachment complex comprises at least two anchors. In certain instances, the attachment complex is attached to a support. In certain other instances, the at least two anchors in the attachment complex further comprises a topological tether. In yet certain other instances, the topological tether is an antibody and the at least two anchors are for example, each a histidine tag.

[0013] In another embodiment, the attachment complex comprises a topological tether. In certain instances, the topological tether comprises an antibody. In yet another embodiment, the topological tether is attached to the at least one anchor via a complementary binding pair.

In a further embodiment, the topological tether is attached to the at least two anchors via at least two complementary binding pairs.

[0014] In another embodiment, the at least one anchor comprises an at least one amino acid or an epitope for attachment. In certain instances, the at least one amino acid is selected from the group of a cysteine, a phenylalanine derivative or a histidine. In certain other instances, the histidine is selected from the group of a histidine tag, a histidine patch or a polyhistidine sequence.

[0015] In yet another embodiment, the at least one anchor is attached to a support. In certain instances, the at least one anchor entraps the target nucleic acid. In a further embodiment, the target nucleic acid is a circular DNA. In certain instances, the circular DNA is sequenced by strand displacement synthesis.

[0016] In another embodiment, the polymerase is selected from a Family A polymerase and a Family B polymerase. In certain instances, the Family A polymerase is selected from the group of Klenow, Taq, and T7 polymerase. In certain other instances, the Family B polymerase is selected from the group of a Terminator polymerase, phi29, RB-69 and T4 polymerase. In yet another embodiment, the polymerase-nucleic acid complex is an array of polymerase-nucleic acid complexes attached to a support. In certain instances, the plurality of members of the array of polymerase-nucleic acid complexes is randomly attached to the support. In certain other instances, the plurality of members of the array of polymerase-nucleic acid complexes is uniformly attached to the support.

[0017] In a further embodiment, the processivity index is at least 0.5. In certain instances, the processivity index is at least 0.8. In certain other instances, the processivity index is 1.

[0018] In another aspect, the present invention provides a method for detecting incorporation of at least one NTP into a single primer nucleic acid molecule, the method comprising:

- i. immobilizing onto a support a polymerase nucleic acid complex comprising a target nucleic acid, a primer nucleic acid which complements a region of the target nucleic acid, and at least one nucleic acid polymerase;
- ii. contacting said immobilized complex with at least one type of labeled nucleotide triphosphate [NTP], wherein each NTP is labeled with a detectable label, and
- iii. detecting the incorporation of the at least one type of labeled NTP into a single molecule of the primer, while the at least one type of labeled NTP is in contact with

the immobilized complex, by detecting the label of the NTP while the at least one type of labeled NTP is in contact with the polymerase nucleic acid complex.

[0019] In one embodiment, the polymerase nucleic acid complex is contacted with a single type of labeled NTP. In another embodiment, the polymerase nucleic acid complex is contacted with at least two different types of NTPs, and wherein each type of NTP is uniquely labeled. In yet another embodiment, the polymerase nucleic acid complex is contacted with at least four different types of NTPs, and wherein each type of NTP is uniquely labeled. In a further embodiment, the NTPs are labeled on the γ -phosphate. In certain instances, the NTPs are labeled on the γ -phosphate with a fluorescent label.

[0020] In another embodiment, detecting the incorporation of the at least one type of labeled NTP into a single molecule of the primer comprises detecting a unique signal from the labeled NTP using a system or device selected from the group of an optical reader, a high-efficiency photon detection system, a photodiode, a camera, a charge couple device, an intensified charge couple device, a near-field scanning microscope, a far-field confocal microscope, a microscope that detects wide-field epi-illumination, evanescent wave excitation and a total internal reflection fluorescence microscope. In yet another embodiment, the label of the NTP is detected using a method comprising a four color evanescent wave excitation device. In a further embodiment, detecting the incorporation of the at least one type of labeled NTP into a single molecule of the primer is carried out by a mechanism selected from the group of fluorescence resonance energy transfer, an electron transfer mechanism, an excited-state lifetime mechanism and a ground-state complex quenching mechanism.

[0021] In yet another embodiment, detecting the incorporation of the at least one type of labeled NTP into a single molecule of the primer comprises measuring a residence time of a labeled NTP in the polymerase nucleic acid complex. In certain instances, the residence time of an NTP that is incorporated into the primer nucleic acid is at least about 100 times longer to about 10,000 times longer than the residence time of an NTP that is not incorporated. In certain other instances, the residence time of an NTP that is incorporated into the primer nucleic acid is at least about 200 times longer to about 500 times longer than the residence time of an NTP that is not incorporated. In yet certain other instances, the residence time of an NTP that is incorporated into the primer nucleic acid is about 1.0 milliseconds to about

100 milliseconds. In further instances, the residence time of an NTP that is incorporated into the primer nucleic acid is about 2.0 milliseconds to about 10.0 milliseconds.

[0022] In another embodiment, the method of the present invention further comprises the step of genotyping the target nucleic acid by determining the identity of at least one NTP that is incorporated into a single molecule of the primer. In yet another embodiment, the method of the present invention further comprises sequencing the target nucleic acid by determining the identity and sequence of incorporation of NTPs that are incorporated into a single molecule of the primer.

[0023] In a further embodiment, the detection is a sequential detection of the identities of more than one uniquely labeled dNTPs that are sequentially incorporated into the primer, wherein the sequential detection yields the sequence of region of the target DNA that is downstream of the elongating end of the primer. In another embodiment, the polymerase-nucleic acid complex comprises a target nucleic acid and a nucleic acid polymerase, wherein the polymerase has an attachment complex comprising at least one anchor, which irreversibly associates the target nucleic acid with the polymerase for increasing the processivity index.

[0024] These and other objects and advantages will become more apparent when read with the accompanying detailed description and drawings that follow.

DESCRIPTION OF THE DRAWINGS

[0025] Figure 1 illustrates various features of a polymerase-nucleic acid complex of the present invention.

[0026] Figure 2 illustrates an anchor embodiment of the present invention.

[0027] Figure 3 illustrates a nucleic acid sample preparation of the present invention.

[0028] Figure 4 illustrates a nucleic acid sample preparation of the present invention.

[0029] Figure 5 illustrates a nucleic acid sample preparation of the present invention.

[0030] Figure 6 illustrates a single molecule isolation embodiment of the present invention.

[0031] Figure 7 illustrates a single molecule bound to a cover slip.

[0032] Figure 8 illustrates a multiple sequencing embodiment of the present invention.

[0033] Figure 9A-C illustrates a synthetic scheme of a compound useful in the present invention.

[0034] Figure 10 illustrates a schematic view of a setup for a residence-time detector.

[0035] Figure 11 illustrates a computer simulation of incorporation events detected above a
5 signal energy threshold of 2500. The experimental parameters are summarized in Table III.

[0036] Figure 12 illustrates a computer simulation of background incorporation using the same experimental parameters (summarized in Table III) used in Figure 11.

DETAILED DESCRIPTION OF THE INVENTION

10

I. Polymerase-Nucleic acid Complex

[0037] In one embodiment, the present invention provides a polymerase-nucleic acid complex (PNAC), comprising: a target nucleic acid and a nucleic acid polymerase, wherein the polymerase has an attachment complex comprising at least one anchor, which at least one
15 anchor irreversibly associates the target nucleic acid with the polymerase to increase the processivity index. As used herein, the term "processivity index" means the number of nucleotides incorporated before the polymerase dissociates from the DNA. Processivity refers to the ability of the enzyme to catalyze many different reactions without releasing its substrate. That is, the number of phosphodiester bonds formed using the present invention is
20 greatly increased as the substrate is associated with polymerase via an anchor.

[0038] In one embodiment, the processivity index is defined as the number of nucleotides sequenced divided by the number of nucleotides in the template. For example, if the template is 10,000 bases long, and the PNAC sequences 9000 bases, the index is 0.90. Using the PNACs and methods of the present invention, the index is preferably between at least 0.5 to
25 about 1. More preferably, the index is about at least 0.80 to about 1, such as at least 0.80, or at least 0.85, or at least 0.90, or at least 0.95, or 1.0.

[0039] Using the PNACs of the present invention, because the target is irreversibly associated with the polymerase, the number of nucleotides added can be from about 20 to about 100,000, such as about 1000 to about 30,000, such as about 5000 to about 20,000.

[0040] FIG. 1 A-D are examples of polymerase nucleic acid complexes (PNACs) of the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art will recognize other variations, modifications, and alternatives.

- 5 [0041] The polymerase-nucleic complex comprises at least one anchor. In certain aspects, the PNAC will further comprise a primer, which complements a region of the target nucleic acid. As shown in Figure 1A, the polymerase 101 can have at least one anchor 130 such anchor comprising for example, an amino acid, an epitope, a modified amino acid and the like, for attaching a topological tether. The amino acid *i.e.*, anchor can be for example, a
- 10 cysteine or a histidine. In certain aspects, the polymerase nucleic acid complex, wherein the nucleic acid 120 is preferably within the active site, comprises at least two anchors. Suitable anchors of the present invention include, but are not limited to, an amino acid, a modified amino acid, a peptide, a histidine tag, a histidine patch, an epitope, and the like. In certain instances, the at least one anchor entraps the target nucleic acid such as by folding back on
- 15 itself. In other instances, the anchors of the present invention are useful for also attaching a topological tether to the polymerase, or for example, attaching the PNAC to a substrate. In other embodiments, the anchor affixes the PNAC to a support, with or without a topological tether. In certain other embodiments, the polymerase-nucleic complex comprises a
- 20 topological tether bound to at least two anchors.
- [0042] As shown in Figure 1B, an anchor 130 can further comprise other functionalities such as a first member 135 of a first binding pair. A second anchor 140 has a first member 145 of a second binding pair. As shown in Figure 1C, in certain instances, a topological
- 25 tether is formed when the first members 135, 145 are joined by a common member 148. Alternatively, a topological tether can be formed when the first members 135, 145 are each joined directly to a support (not shown). A topological tether and at least one anchor can attach via complementary binding pairs. Alternatively, the anchors can attach directly to a
- 30 substrate without the use of a tether (for example, histidine patches as anchors bound directed to a Ni surface). Suitable complementary binding pairs include, but are not limited to, any haptenic or antigenic compound in combination with a corresponding antibody or binding portion or fragment thereof, nonimmunological binding pairs, receptor-receptor agonist or antagonist, IgG-protein A, lectin-carbohydrate, enzyme-enzyme cofactor, enzyme-enzyme-inhibitor, and complementary polynucleotide pairs capable of forming nucleic acid duplexes.
- [0043] Exemplary complementary binding pairs include, but are not limited to, digoxigenin and anti-digoxigenin, fluorescein and anti-fluorescein, dinitrophenol and anti-dinitrophenol,

bromodeoxyuridine and anti-bromodeoxyuridine, mouse immunoglobulin and goat anti-mouse immunoglobulin, biotin-avidin, biotin-streptavidin, thyroxine and cortisol, histidine patch and Ni-NTA and acetylcholine and receptor-acetylcholine. In certain aspects, the anchor comprises at least one amino acid or an epitope for attaching the topological tether.

5 [0044] As discussed, in certain instances, anchors can comprise an amino acids capable of modification for attachment to a binding member, a tether, a support, and combinations thereof. In one embodiment, a topological tether can attach to two anchors, without intervening binding pairs.

[0045] In one aspect, the anchor comprises a biotin moiety. For example, biotin-X
10 nitrilotriacetic acid can be used to covalently attach the biotin moiety to a protein having a free amino group. In turn, this biotin anchor can attach to a streptavidin or a neutravidin binding member, or alternatively, directly to a streptavidin or a neutravidin support.

[0046] In another aspect, the topological tether comprises an antibody. In certain
15 embodiments, the topological tether is an antibody that can attach via anchors having complementary binding pairs. For example, the two anchors can be histidine tags, and the tether can be an antibody. In certain aspects, the polymerase-nucleic complex comprises a topological tether anchored to a solid support 150 (see, Figure 1D).

[0047] In certain aspects, the polymerase-nucleic acid attachment complex can be attached
20 to the substrate by providing an anchor such as a polyhistidine tag, that binds to metal. Other conventional means for attachment employ binding pairs. Alternatively, covalent crosslinking agents can be employed such as reagents capable of forming disulfide (S-S), glycol (-CH(OH)-CH(OH)-), azo (-N=N-), sulfone (-S(=O)₂-), ester (-C(=O)-O-), or amide (-C(=O)-N-) bridges. The covalent bond is for example, an amide, a secondary or tertiary amine, a carbamate, an ester, an ether, an oxime, a phosphate ester, a sulfonamide, a
25 thioether, a thiourea, or a urea.

[0048] Selected examples of reactive functionalities useful for the attaching an anchor to the polymerase, a tether to the anchor, or the PNAC to the substrate are shown in Table I, wherein the bond results from such a reaction. Those of skill in the art will know of other bonds suitable for use in the present invention.

TABLE I

Reactive functionality	Complementary group	The resulting bond
activated esters	amines/anilines	carboxamides
acrylamides	thiols	thioethers

acyl azides	amines/anilines	carboxamides
acyl halides	amines/anilines	carboxamides
acyl halides	alcohols/phenols	esters
acyl nitriles	alcohols/phenols	esters
acyl nitriles	amines/anilines	carboxamides
aldehydes	amines/anilines	imines
aldehydes or ketones	hydrazines	hydrazones
aldehydes or ketones	hydroxylamines	oximes
alkyl halides	amines/anilines	alkyl amines
alkyl halides	carboxylic acids	esters
alkyl halides	thiols	thioethers
alkyl halides	alcohols/phenols	ethers
alkyl sulfonates	thiols	thioethers
alkyl sulfonates	carboxylic acids	esters
alkyl sulfonates	alcohols/phenols	ethers
anhydrides	alcohols/phenols	esters
anhydrides	amines/anilines	carboxamides/imides
aryl halides	thiols	thiophenols
aryl halides	amines	aryl amines
aziridines	thiols	thioethers
boronates	glycols	boronate esters
carboxylic acids	amines/anilines	carboxamides
carboxylic acids	alcohols	esters
carboxylic acids	hydrazines	hydrazides
carbodiimides	carboxylic acids	N-acylureas or anhydrides
dialkyl azides	carboxylic acids	esters
epoxides	thiols (amines)	thioethers (alkyl amines)
epoxides	carboxylic acids	esters
haloacetamides	thiols	thioethers
haloplatinate	amino	platinum complex
haloplatinate	heterocycle	platinum complex
halotriazines	amines/anilines	aminotriazines
halotriazines	alcohols/phenols	triazinyl ethers
imido esters	amines/anilines	amidines
isocyanates	amines/anilines	ureas
isocyanates	alcohols/phenols	urethanes
isothiocyanates	amines/anilines	thioureas
maleimides	thiols	thioethers
phosphoramidites	alcohols	phosphite esters
silyl halides	alcohols	silyl ethers
sulfonate esters	amines/anilines	alkyl amines
sulfonyl halides	amines/anilines	sulfonamides

[0049] In certain aspects, the polymerase can be covalently attached to a support (e.g., coverslip, metal surface, and the like), wherein the polymerase is labeled *in vivo* with a modified amino acid such as for example, a benzaldehyde derivative of phenylalanine. In

one example, the benzaldehyde derivative of phenylalanine is p-acetyl-L-phenylalanine, which can be labeled at specific position(s) in the polymerase. This can be accomplished using organisms (*e.g.*, *E. coli*, yeast) engineered to have an augmented 21-amino acid genetic code capable of inserting p-acetyl-L-phenylalanine at specific codons (*see*, Lei Wang,

5 Zhiwen Zhang, Ansgar Brock, Peter G. Schultz (2003) *Proc Natl Acad Sci USA* 100:56-61).

In one aspect, the polymerase gene of the present invention is engineered to have the appropriate codon or codons at the desired anchor positions, and the corresponding polymerase protein is expressed in the 21-amino acid organism. The expressed polymerase is then purified, mixed with the template DNA, and the resulting PNACs are contacted to a support derivatized with a hydrazine, hydrazone, and the like (*e.g.*, SANH from Solulink Inc). Alternatively, a chemical functionality equivalent to p-acetyl-L-phenylalanine can be attached to the protein at specific or unspecific positions by conjugating SFB (Solulink Inc) to lysine amino acids on the protein. The functionalized protein is attached to the support as above.

15 [0050] FIG. 2 shows a structural model of a PNAC comprising a 9 Degrees North DNA polymerase (parent of Terminator polymerase) 202 and a circular primed DNA template 200. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art will recognize other variations, modifications, and alternatives. The polymerase 202 comprises anchors 203 and 205 inserted at Terminator amino acid positions K53 and K229, respectively. The anchors are identical in amino acid sequence (LLSKKRSLCCXCTVIVYVTDT), wherein the anchor comprises amino acid pa-Phe, which is indicated by "X" in the sequence and by white diamonds 204, 206. The pa-Phe amino acids 204, 206 are shown attached to the support 207. The circular DNA template 200 is hybridized to a primer 201. The 5'-end of the primer is indicated 201 and the 3'-end of the primer is hidden in the DNA binding cleft of the protein 202. The structural model is 1QHT.pdb in the protein database at <http://www.rcsb.org/pdb/>.

[0051] As discussed, the Terminator DNA polymerase can be modified by inserting a 20-amino acid anchor at position K53 and a 20-amino acid anchor at position K229 in the Terminator gene. These two positions straddle the DNA binding cleft as shown in Fig 2.

30 As shown therein, each 20-amino acid anchor is engineered to contain at least one p-acetyl-L-phenylalanine (pa-Phe) amino acid near the middle of the anchor (Fig. 2). The engineered protein is then purified. In one embodiment, to make polymerase nucleic acid complexes, the purified Terminator protein is mixed with a primed single stranded circular DNA template

and the mixture is contacted with a support derivatized with hydrazine or hydrazone linkers (Solulink Inc). Optionally, the template DNA contains at least one dUTP base positioned 4-5 bases from the 3'-end of the primer in order to stabilize the polymerase-DNA complex as described (*see*, Mark Fogg, Laurence Pearl, Bernard Connolly (2002) *Nature Structural Biology* 9:922-927). The polymerase-DNA complex attaches to the support by bond formation between the pa-Phe on the protein and the hydrazine or hydrazone linker on the support. Optionally, the kinetics of bond formation can be increased by concentrating polymerase-DNA complexes on the support surface using an energy field (*e.g.*, electric field, pressure field, magnetic field, and the like). Once the PNAC has formed on the support, the circular DNA is irreversibly associated with the polymerase as shown in Fig. 2.

A. Polymerases

[0052] The polymerases suitable for use in the present invention preferably have a fidelity (incorporation accuracy) of at least 99%. In addition, the processivity of the polymerase should be at least 20 nucleotides, prior to immobilization. Although the polymerase selected for use in this invention is not critical, preferred polymerases are able to tolerate labels on the γ -phosphate of the NTP.

[0053] In certain aspects, the polymerases useful in the present invention are selected from the A family polymerases or the B family polymerases. DNA-dependent DNA polymerases have been grouped into families, including A, B, X, and others on the basis of sequence similarities. Members of family A, which includes bacterial and bacteriophage polymerases, share significant similarity to *E.coli* polymerase I; hence family A is also known as the pol I family. The bacterial polymerases also contain an exonuclease activity, which is coded for in the N-terminal portion. Family A polymerases include for example, Klenow, Taq, and T7 polymerases. Family B polymerases include for example, the Terminator polymerase, phi29, RB-69 and T4 polymerases.

[0054] In certain instances, suitable DNA polymerases can be modified for use in the present invention. These polymerases include, but are not limited to, DNA polymerases from organisms such as *Thermus flavus*, *Pyrococcus furiosus*, *Thermotoga neapolitana*, *Thermococcus litoralis*, *Sulfolobus solfataricus*, *Thermatoga maritima*, *E. coli* phage T5, and *E. coli* phage T4. The DNA polymerases may be thermostable or not thermostable.

[0055] In other embodiments, the polymerases include T7 DNA polymerase, T5 DNA polymerase, HIV reverse transcriptase, *E. coli* DNA pol I, T4 DNA polymerase, T7 RNA polymerase, Taq DNA polymerase and *E. coli* RNA polymerase. In certain instances, exonuclease-defective versions of these polymerases are preferred. The efficiency with which γ -labeled NTPs are incorporated may vary between polymerases; HIV-1 RT and *E. coli* RNA polymerase reportedly readily incorporate γ -labeled nucleotide. The polymerase can also be a T7 polymerase. T7 polymerase has a known 3D structure and is known to be processive. In order to operate in a strand-displacement mode, the polymerase requires a complex of three proteins: T7 polymerase + thioredoxin + primase (Chowdhury *et al.* *PNAS* 97:12469). In other embodiments, the polymerases can also be HIV RT and DNA Polymerase I.

B. Sources of target nucleic acid.

[0056] The identity and source of the template and primer nucleic acid ("NA") is generally not critical, although particular NAs are needed for specific applications. NA used in the present invention can be isolated from natural sources, obtained from such sources such as ATCC, GenBank libraries or commercial vendors, or prepared by synthetic methods. It can be mRNA, ribosomal RNA, genomic DNA or cDNA, an oligonucleotide, which can be either isolated from a natural source or synthesized by known methods. When the target (*i.e.*, template) NA is from a biological source, there are a variety of known procedures for extracting nucleic acid and optionally amplified to a concentration convenient for genotyping or sequence work. Nucleic acid can be obtained from any living cell of a person, animal or plant. Humans, pathogenic microbes and viruses are particularly interesting sources.

[0057] Nucleic acid amplification methods are also known and can be used to generate nucleic acid templates for sequencing. Preferably, the amplification is carried out by polymerase chain reaction (PCR) (U.S. Pat. Nos. 4,683,202, 4,683,195 and 4,889,818; Gyllenstein *et al.*, 1988, *Proc. Natl. Acad. Sci. USA* 85: 7652-7656; Ochman *et al.*, 1988, *Genetics* 120: 621-623; Loh *et al.*, 1989, *Science* 243: 217-220; Innis *et al.*, 1990, PCR PROTOCOLS, Academic Press, Inc., San Diego, Calif.). Other amplification methods known in the art can be used, including but not limited to ligase chain reaction, use of Q- beta replicase, or methods listed in Kricka *et al.*, 1995, MOLECULAR PROBING, BLOTTING, AND SEQUENCING, Chap. 1 and Table IX, Academic Press, New York.

[0058] Any NA used in the invention can also be synthesized by a variety of solution or solid phase methods. Detailed descriptions of the procedures for solid phase synthesis of nucleic acids by phosphite-triester, phosphotriester, and H-phosphonate chemistries are widely available. See, for example, Itakura, U.S. Pat. No. 4,401,796; Caruthers, *et al.*, U.S. Pat. Nos. 4,458,066 and 4,500,707; Beaucage, *et al.*, Tetrahedron Lett., **22**:1859-1862 (1981); Matteucci, *et al.*, J. Am. Chem. Soc., **103**:3185-3191 (1981); Caruthers, *et al.*, Genetic Engineering, **4**:1-17 (1982); Jones, chapter 2, Atkinson, *et al.*, chapter 3, and Sproat, *et al.*, chapter 4, in *Oligonucleotide Synthesis: A Practical Approach*, Gait (ed.), IRL Press, Washington D.C. (1984); Froehler, *et al.*, Tetrahedron Lett., **27**:469-472 (1986); Froehler, *et al.*, Nucleic Acids Res., **14**:5399-5407 (1986); Sinha, *et al.* Tetrahedron Lett., **24**:5843-5846 (1983); and Sinha, *et al.*, Nucl. Acids Res., **12**:4539-4557 (1984) which are incorporated herein by reference.

[0059] In one preferred embodiment, the target nucleic acid is circular DNA. In one aspect, the circular DNA is sequenced by strand displacement synthesis. As is shown in FIG. 3, randomly-sheared fragments of genomic DNA are purified from a sample organism. The DNA 300 is then treated with for example, T4 DNA polymerase, to generate blunt ends and a single "A" nucleotide is added to the 3'-ends with for example, Taq DNA polymerase, and dATP. A mixture of two double-stranded oligonucleotide adaptors 301 and 302 (each with a "T" nucleotide on one 3'-end to complement the "A" nucleotide on the randomly-sheared fragment) is ligated to the DNA fragments 300 with T4 DNA ligase, wherein the first adaptor 301 is 5'-biotinylated on one strand and the second adaptor 302 is not biotinylated. Whereas the adaptors attach with equal probability to the DNA fragment ends, about half of the ligated DNA molecules will have one biotinylated adaptor and one non-biotinylated adaptor, one quarter will have two biotinylated adaptors, and one quarter will have two non-biotinylated adaptors as shown in FIG. 3. The desired ligated DNA fragment types, having one biotinylated and one non-biotinylated adaptor, are purified after ligation using gel electrophoresis and streptavidin-coated magnetic beads as follows.

[0060] After ligation, DNA fragments in the size range of about 17-23 kb are purified by gel electrophoresis. As shown in FIG. 4, the purified fragments are bound to streptavidin-coated magnetic beads (Dynal). After binding, the beads are washed to remove unbound DNA. Then the bound DNA is denatured at alkaline pH and the unbiotinylated strands 401 are eluted and the DNA still bound to the beads is discarded. As shown in FIG. 5, the eluted

strands are circularized by hybridizing to a primer oligonucleotide complementary to both adaptors and ligating the two ends of the eluted strand.

C. Immobilization of the PNACs

[0061] In certain embodiments, the PNAC arrays of the present invention are immobilized on a support. Preferably, the support (*e.g.*, solid support) comprises a bioreactive moiety or bioadhesive layer. The support can be for example, glass, silica, plastic or any other conventionally material that will not create significant noise or background for the detection methods. The bioadhesive layer can be an ionic adsorbent material such as gold, nickel, or copper, protein-adsorbing plastics such as polystyrene (US Pat. No. 5,858,801), or a covalent reactant such as a thiol group.

[0062] The PNAC arrays of the present invention can be immobilized on a support in a random fashion (*e.g.*, random X or Y position coordinates), uniform fashion (*e.g.*, regularly spaced X or Y position coordinates) or a combination thereof. As is shown in FIG. 6, in one aspect, the PNAC are isolated into single molecule configuration. This single molecule isolation enables efficient attachment of the PNACs to the support. In addition, it allows for efficient single molecule sequencing. Advantageously, the present invention provides single PNACs attached so as to be optically resolvable from their nearest neighbor PNACs. Thus, the PNACs can be analyzed individually without interference from overlapping optical signals from neighboring PNACs. In the present invention, many individual optically resolved PNACs can be sequenced simultaneously.

[0063] FIG. 7 is an example of a randomly associated array of PNACs immobilized on a neutravidin-coated slide. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art will recognize other variations, modifications, and alternatives. As shown therein, PNACs are attached or immobilized to a neutravidin-coated slide via an anchor having for example, the first member of a binding pair, wherein the anchor comprises a biotin moiety. In operation, multiple sites can be sequenced with ease.

[0064] In yet another example, the PNACs can be attached to the bioadhesive pattern by providing a polyhistidine tag on the polymerase that binds to metal bioadhesive patterns. To create a patterned or random array of a bioadhesive layer, an electron-sensitive polymer such as polymethyl methacrylate (PMMA) coated onto the support is etched in any desired pattern

using an electron beam followed by development to remove the sensitized polymer. The holes in the polymer are then coated with a metal such as nickel, and the polymer is removed with a solvent, leaving a pattern of metal posts on the substrate. This method of electron beam lithography provides the very high spatial resolution and small feature size required to immobilize just one molecule at each point in the patterned array. An alternate means for creating high-resolution patterned arrays is atomic force microscopy. A third means is X-ray lithography.

[0065] Other conventional means for attachment employ homobifunctional and heterobifunctional crosslinking reagents. Homobifunctional reagents carry two identical functional groups, whereas heterobifunctional reagents contain two dissimilar functional groups to link the biologics to the bioadhesive. A vast majority of the heterobifunctional cross-linking agents contain a primary amine-reactive group and a thiol-reactive group. Covalent crosslinking agents are selected from reagents capable of forming disulfide (S-S), glycol (-CH(OH)-CH(OH)-), azo (-N=N-), sulfone (-S(=O₂-), ester (-C(=O)-O-), or amide (-C(=O)-N-) bridges.

[0066] A bioresist layer may be placed or superimposed upon the bioadhesive layer either before or after attachment of the biologic to the bioadhesive layer. The bioresist layer is any material that does not bind the biologic. Examples include bovine serum albumin, neutravidin, gelatin, lysozyme, octoxynol, polysorbate 20 (polyethenesorbitan monolaurate) and polyethylene oxide containing block copolymers and surfactants (US Pat. No. 5,858,801). Deposition of the layers is done by conventional means, including spraying, immersion and evaporative deposition (metals).

II. Methods

[0067] The present invention provides *inter alia*, methods to detect incorporation of a detectably labeled nucleotide triphosphate ("NTP") onto the growing end of a primer nucleic acid molecule. The method is used, for example, to genotype and sequence a nucleic acid. In turn, the sequence identification can be used to identify metabolic differences in patient groups based upon genetic polymorphism to provide improved dosing regimens, enhancing drug efficacy and safety. Further, understanding the genetic basis of disease in animal and plants will help engineer disease resistant animals & crops as well as enhance desirable characteristics.

[0068] In a preferred embodiment, the methods described herein detect the "residence time" of an individual fluorogenic NTP molecule on a PNAC preferably comprised of at least one RNA or DNA dependent polymerase, a single target nucleic acid template, and a single primer nucleic acid. The NTPs are preferably labeled with a fluorescent dye, which is preferably attached to the γ -phosphate. As shown in FIG. 8, as the polymerase moves along the target nucleic acid, the nucleotide sequence is read by identifying the order and identity of incorporated NTPs. In one embodiment, all the NTPs have the same label, but each class of labeled NTPs is sequentially added to the complex; the incorporated NTP corresponds to the particular class that is being infused.

[0069] In another embodiment, at least two classes of NTP are used, or at least three classes of NTP are used, or at least four classes of NTP are used each of which is uniquely labeled. The identity of the NTP incorporated during a particular incorporation event is determined by detecting the unique label of the incorporated NTP, based on the residence time or the time-averaged intensity of the labeled NTP in contact with the PNAC.

[0070] The NTPs can optionally include a fluorescence quencher attached to either the base sugar, dye, polymerase, or combinations thereof, which quenches the fluorescence of the fluorescent dye while the NTP (γ -label) is free in solution. The fluorescence associated with the immobilized complex is detected. Upon interaction with the complex, the fluorescence of the labeled NTP changes (e.g., increases), as the conformation of the NTP is altered by interaction with the complex, and/or as the PPi is cleaved prior to being released into the medium. The optical properties of the pyrophosphate-dye moiety change, either by conformational changes of the NTP or cleavage of the PPi, which in turn facilitates detection of the fluorescent dye.

A. Labeling of NTPs

1. Attachment of a γ -Phosphate Fluorophore

[0071] The methods of the present invention involve detecting and identifying individual detectably labeled NTP molecules as a polymerase incorporates them into a single nucleic acid molecule. Suitable nucleobases include, but are not limited to, adenine, guanine, cytosine, uracil, thymine, deazaadenine and deazaguanosine. In certain preferred embodiments, a fluorophore is attached to the γ -phosphate of the NTP by known methods.

[0072] The fluorophore may be any known fluorophore including, but not limited to, the following:

TABLE II

<u>FLUOROPHORE</u>	<u>Absorbance/Emission</u>
Rho123	507/529
R6G	528/551
BODIPY 576/589	576/589
BODIPY TR	588/616
Nile Blue	627/660
BODIPY 650/665	650/665
Sulfo-IRD700	680/705
NN382	778/806
Tetramethylrhodamine	550
Rodamine X	575
Cy3 TM	550
Cy5 TM	650
Cy7 TM	750

5 [0073] There is a great deal of practical guidance available in the literature for providing an exhaustive list of fluorescent and chromogenic molecules and their relevant optical properties (see, for example, Berlman, *Handbook of Fluorescence Spectra of Aromatic Molecules*, 2nd Edition (Academic Press, New York, 1971); Griffiths, *Colour and Constitution of Organic Molecules* (Academic Press, New York, 1976); Bishop, Ed., *Indicators* (Pergamon Press, Oxford, 1972); Haugland, *Handbook of Fluorescent Probes and Research Chemicals* (Molecular Probes, Eugene, 1992) Pringsheim, *Fluorescence and Phosphorescence* (Interscience Publishers, New York, 1949); and the like. Further, there is extensive guidance in the literature for derivatizing fluorophore and quencher molecules for covalent attachment via common reactive groups that can be added to a nucleotide, as exemplified by the following references: Haugland (*supra*); Ullman *et al.*, U.S. Pat. No. 3,996,345; Khanna *et al.*, U.S. Pat. No. 4,351,760.

[0074] There are many linking moieties and methodologies for attaching fluorophore or quencher moieties to nucleotides, as exemplified by the following references: Eckstein, editor, *OLIGONUCLEOTIDES AND ANALOGUES: A PRACTICAL APPROACH* (IRL Press, Oxford, 1991); Zuckerman *et al.*, *Nucleic Acids Research*, 15: 5305-5321 (1987) (3' thiol group on oligonucleotide); Sharma *et al.*, *Nucleic Acids Research*, 19: 3019 (1991) (3' sulfhydryl); Giusti *et al.*, *PCR Methods and Applications*, 2: 223-227 (1993); Fung *et al.*, U.S. Pat. No. 4,757,141 (5' phosphoamino group via AminolinkTM. II available from Applied Biosystems,

Foster City, Calif.); Stabinsky, U.S. Pat. No. 4,739,044 (3' aminoalkylphosphoryl group); Agrawal *et al.*, *Tetrahedron Letters*, **31**: 1543-1546 (1990) (attachment via phosphoramidate linkages); Sproat *et al.*, *Nucleic Acids Research*, **15**: 4837 (1987) (5' mercapto group); Nelson *et al.*, *Nucleic Acids Research*, **17**: 7187-7194 (1989) (3' amino group); and the like.

- 5 [0075] In general, nucleoside labeling can be accomplished using any of a large number of known nucleoside labeling techniques using known linkages, linking groups, and associated complementary functionalities. The linkage linking the quencher moiety and nucleoside should be compatible with relevant polymerases and not quench the fluorescence of the fluorophore moiety.
- 10 [0076] Suitable dyes operating on the principle of fluorescence energy transfer (FET) include, but are not limited to, 4-acetamido-4'-isothiocyanatostilbene-2,2'-disulfonic acid; acridine and derivatives: acridine, acridine isothiocyanate; 5-(2'-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS); 4-amino-N-[3-vinylsulfonyl]phenyl]naphthalimide-3,5 disulfonate; N-(4-anilino-1-naphthyl)maleimide;
- 15 anthranilamide; BODIPY; Brilliant Yellow; coumarin and derivatives: coumarin, 7-amino-4-methylcoumarin (AMC, Coumarin 120), 7-amino-4-trifluoromethylcoumarin (Coumarin 151); cyanine dyes; cyanosine; 4',6-diaminidino-2-phenylindole (DAPI); 5', 5"-dibromopyrogallol-sulfonaphthalein (Bromopyrogallol Red); 7-diethylamino-3-(4'-isothiocyanatophenyl)-4-methylcoumarin; diethylenetriamine pentaacetate; 4,4'-
- 20 diisothiocyanatodihydro-stilbene-2,2'-disulfonic acid; 4,4'-diisothiocyanatostilbene-2,2'-disulfonic acid; 5-[dimethylamino]naphthalene-1-sulfonyl chloride (DNS, dansylchloride); 4-dimethylaminophenylazophenyl-4'-isothiocyanate (DABITC); eosin and derivatives: eosin, eosin isothiocyanate, erythrosin and derivatives: erythrosin B, erythrosin, isothiocyanate; ethidium; fluorescein and derivatives: 5-carboxyfluorescein (FAM), 5-(4,6-dichlorotriazin-2-yl)aminofluorescein (DTAF), 2',7'-dimethoxy-4'5'-dichloro-6-carboxyfluorescein (JOE),
- 25 fluorescein, fluorescein isothiocyanate, QFITC, (XRITC); fluorescamine; IR144; IR1446; Malachite Green isothiocyanate; 4-methylumbelliferoneortho cresolphthalein; nitrotyrosine; pararosaniline; Phenol Red; B-phycoerythrin; o-phthalaldehyde; pyrene and derivatives: pyrene, pyrene butyrate, succinimidyl 1-pyrene; butyrate quantum dots; Reactive Red 4
- 30 (Cibacron™ Brilliant Red 3B-A) rhodamine and derivatives: 6-carboxy-X-rhodamine (ROX), 6-carboxyrhodamine (R6G), lissamine rhodamine B sulfonyl chloride rhodamine (Rhod), rhodamine B, rhodamine 123, rhodamine X isothiocyanate, sulforhodamine B, sulforhodamine 101, sulfonyl chloride derivative of sulforhodamine 101 (Texas Red);

N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA); tetramethyl rhodamine; tetramethyl rhodamine isothiocyanate (TRITC); riboflavin; rosolic acid; terbium chelate derivatives; Cy3; Cy5; Cy5.5; Cy7; IRD 700; IRD 800; La Jolla Blue; phthalo cyanine; and naphthalo cyanine.

[0077] In certain embodiments, certain visible and near IR dyes are known to be sufficiently fluorescent and photostable to be detected as single molecules. In this aspect the visible dye, BODIPY R6G (525/545), and a larger dye, LI-COR's near-infrared dye, IRD-38 (780/810) can be detected with single-molecule sensitivity and are used to practice the present invention.

2. Exemplary labeled nucleotides

(i) dATP-PEG-TAMRA

(a) Deprotection of BOC-PEG8-amine (2)

[0078] Turning now to FIG. 9 A, BOC-PEG8-amine (1) (1g), purchased from PolyPure, is added to a 50% trifluoroacetic acid/chloroform solution (20 mL). The mixture is stirred at room temperature for several hours, and then concentrated down *in vacuo* to a light orange viscous liquid.

(b) Gamma labeled dATP (4) with PEG-diamine (2)

[0079] With respect to FIG. 9B, dATP (3) (1eq., 6.3×10^{-3} mmol, 3.4 mg, 79 mM; Sigma) and EDC (2.5×10^{-1} mmol, 48.8mg, 6.5M; Aldrich) are added together in 500 mM MES at pH 5.8. The mixture is allowed to react at room temperature for 10 min. and is then added to the PEG-diamine solution (2) (10 eq., 6.3×10^{-2} mmol, 37.5 mg, 31mM). The pH is adjusted to 5.8-6 using 5M KOH before adding to the nucleotide. The mixture is allowed to react at room temperature for a minimum of 3 hours. The product is first purified on a HiPrep DEAE column (Amersham) using buffer A (10mM phosphate + 20%ACN) and buffer B (Buffer A in 1M NaCl) by holding in buffer A for 10min and then applying a 0-100% buffer B gradient for 5 minutes. The free PEG is eluted from the column, and then the nucleotide is eluted and collected. A second purification is performed on an Inerstil 10 μ m C18 column using buffer A (100mM TEAAc, pH 6.6-6.8, 4%ACN) and buffer B (100 mM TEAAc, pH 6.6-6.8, 80% CAN) over a period of 15 min. The product is dried *in vacuo*.

(c) dATP-PEG-TAMRA (6)

[0080] With respect to FIG. 9C, the dATP-PEG-amine (4) product is reconstituted in water and quantitated using UV-VIS. dATP-PEG-amine (9.5×10^{-5} mmol, 5 μ l, 1 eq.), 29 μ l in 50mM carbonate buffer, pH 8, and TAMRA-X SE (5) (1.5eq., 1.4×10^{-4} mmol, 9 μ l of stock dye solution dissolved at a concentration of 10mg/mL in DMF; Molecular Probes) are added together. The reaction proceeds at room temperature for 2 hrs. in the dark. Purification of the product is carried out using a HiPrep DEAE column (Amersham) with buffer A (10mM phosphate + 20%ACN) and buffer B (buffer A in 1M NaCl) by holding in buffer A for 10min and then applying a 0-100% buffer B gradient for 5 minutes. The product is eluted in the void volume. The fractions are collected and concentrated. A second purification step is performed using an Inertsil C18 column with buffer A (100mM TEAAc, pH 6.6-6.8, 4%ACN) and buffer B (100mM TEAAc, pH 6.6-6.8, 80%) by applying a 20-100% buffer B gradient over a period of 15 min. The product is dried *in vacuo*.

[0081] In some embodiments of the present invention, detection of pyrophosphate may involve dequenching, or turning on, a quenched fluorescent dye. Efficient quenching lowers background fluorescence, thus enhancing the signal (unquenched NTP fluorescence) -to-noise (quenched NTP fluorescence) ratio. Incomplete quenching results in a low level fluorescence background from each dye molecule. Additional background fluorescence is contributed by a few of the dye molecules that are fully fluorescent because of accidental (*i.e.*, pyrophosphate-independent) dequenching, for example by breakage of a bond connecting the dye to the quencher moiety. Thus, the background fluorescence has two components: a low-level fluorescence from all dye molecules, referred to herein as "distributed fluorescence background" and full-strength fluorescence from a few molecules, referred to herein as "localized fluorescence background."

[0082] In instances where a multi-labeling scheme is utilized, a wavelength which approximates the mean of the various candidate labels' absorption maxima may be used. Alternatively, multiple excitations may be performed, each using a wavelength corresponding to the absorption maximum of a specific label. Table II lists examples of various types of fluorophores and their corresponding absorption maxima.

B. Miscellaneous reaction reagents.

[0083] The primers (DNA polymerase) or promoters (RNA polymerase) are synthetically made using conventional nucleic acid synthesis technology. The complementary strands of the probes are conveniently synthesized on an automated DNA synthesizer, *e.g.* an Applied

Biosystems, Inc. (Foster City, Calif.) model 392 or 394 DNA/RNA Synthesizer, using standard chemistries, such as phosphoramidite chemistry, *e.g.* disclosed in the following references: Beaucage and Iyer, *Tetrahedron*, 48: 2223-2311 (1992); Molko *et al.*, U.S. Pat. No. 4,980,460; Koster *et al.*, U.S. Pat. No. 4,725,677; Caruthers *et al.*, U.S. Pat. Nos.

5 4,415,732; 4,458,066; and 4,973,679; and the like. Alternative chemistries, *e.g.* resulting in non-natural backbone groups, such as phosphorothioate, phosphoramidate, and the like, may also be employed provided that the resulting oligonucleotides are compatible with the polymerase. They can be ordered commercially from a variety of companies which specialize in custom oligonucleotides.

10 [0084] Primers in combination with polymerases are used to sequence target DNA. Primer length is selected to provide for hybridization to complementary template DNA. The primers will generally be at least 10 bp in length, usually at least between 15 and 30 bp in length. Primers are designed to hybridize to known internal sites on the subject target DNA. Alternatively, the primers can bind to synthetic oligonucleotide adaptors joined to the ends of
15 target DNA by a ligase. Similarly where promoters are used, they can be internal to the target DNA or ligated as adaptors to the ends.

C. Reaction conditions.

[0085] The reaction mixture for the sequencing using the PNACs and methods of the present invention comprises an aqueous buffer medium which is optimized for the particular
20 polymerase. In general, the buffer includes a source of monovalent ions, a source of divalent cations and a buffering agent. Any convenient source of monovalent ions, such as KCl, K-acetate, NH₄-acetate, K-glutamate, NH₄Cl, ammonium sulfate, and the like may be employed, where the amount of monovalent ion source present in the buffer will typically be present in an amount sufficient to provide for a conductivity in a range from about 500 to
25 20,000, usually from about 1000 to 10,000, and more usually from about 3,000 to 6,000 microhms.

[0086] The divalent cation may be magnesium, manganese, zinc and the like, where the cation will typically be magnesium. Any convenient source of magnesium cation may be employed, including MgCl₂, Mg-acetate, and the like. The amount of Mg ion present in the
30 buffer may range from 0.5 to 20 mM, but will preferably range from about 1 to 12 mM, more preferably from 2 to 10 mM and will ideally be about 5 mM.

[0087] Representative buffering agents or salts that may be present in the buffer include Tris, Tricine, HEPES, MOPS and the like, where the amount of buffering agent will typically range from about 5 to 150 mM, usually from about 10 to 100 mM, and more usually from about 20 to 50 mM, where in certain preferred embodiments the buffering agent will be present in an amount sufficient to provide a pH ranging from about 6.0 to 9.5, where most preferred is pH 7.6 at 25° C. Other agents which may be present in the buffer medium include chelating agents, such as EDTA, EGTA and the like.

D. Sample Housing.

[0088] The support is optionally housed in a flow chamber having an inlet and outlet to allow for renewal of reactants which flow past the immobilized moieties. The flow chamber can be made of plastic or glass and should either be open or transparent in the plane viewed by the microscope or optical reader. Electro-osmotic flow requires a fixed charge on the solid support and a voltage gradient (current) passing between two electrodes placed at opposing ends of the solid support. The flow chamber can be divided into multiple channels for separate sequencing. Examples of micro flow chambers exist. For example, Fu *et al.* (*Nat. Biotechnol.* (1999) 17:1109) describe a microfabricated fluorescence-activated cell sorter with 3 μm x 4 μm channels that utilizes electro-osmotic flow for sorting.

E. Detection of fluorophores.

[0089] Various detectors are suitable for use in the present invention. These include, but are not limited to, an optical reader, a high-efficiency photon detection system, a photodiode, a camera, a charge couple device, an intensified charge couple device, a near-field scanning microscope, a far-field confocal microscope, a microscope that detects wide-field epi-illumination, evanescent wave excitation and a total internal reflection fluorescence microscope. In certain aspects, the detection requires the imaging of single molecules in a solution. There are a variety of known ways of achieving this goal, including those described in: Basche *et al.*, eds., 1996, "Single molecule optical detection, imaging, and spectroscopy," Weinheim *et al.*, "Single-molecule spectroscopy," *Ann. Rev. Phys. Chem.* 48: 181-212; Soper *et al.*, "Detection and Identification of Single Molecules in Solution," *J. Opt. Soc. Am. B*, 9(10): 1761-1769, Oct. 1992; Keller *et al.* (1996), *Appl. Spectrosc.* 50: A12-A32; Goodwin *et al.* (1996), *Accounts Chem. Res.* 29: 607-613; Rigler (1995). *J. Biotech.*, 41: 177; Rigler *et al.* Fluorescence Spectroscopy; Wolfbeis O. S., Ed.; Springer, Berlin, 1992, pp 13-

24; Edman *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93: 6710; Schmidt *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 1 93: 2926; Keller *et al.* (1996) *Appl. Spectroscopy* 50: A12.

[0090] A laser source is often used as the excitation source for ultrasensitive measurements but conventional light sources such as rare gas discharge lamps and light emitting diodes (LEDs) are also used. The fluorescence emission can be detected by a photomultiplier tube, photodiode or other light sensor. An array detector such as a charge-coupled device (CCD) detector can be used to image an analyte spatial distribution.

[0091] Raman spectroscopy can be used as a detection method for microchip devices with the advantage of gaining molecular vibrational information. Sensitivity has been increased through surface enhanced Raman spectroscopy (SERS) effects but only at the research level. Electrical or electrochemical detection approaches are also of particular interest for implementation on microchip devices due to the ease of integration onto a microfabricated structure and the potentially high sensitivity that can be attained. The most general approach to electrical quantification is a conductometric measurement, i.e., a measurement of the conductivity of an ionic sample. The presence of an ionized analyte can correspondingly increase the conductivity of a fluid and thus allow quantification. Amperimetric measurements imply the measurement of the current through an electrode at a given electrical potential due to the reduction or oxidation of a molecule at the electrode. Some selectivity can be obtained by controlling the potential of the electrode but it is minimal. Amperimetric detection is a less general technique than conductivity because not all molecules can be reduced or oxidized within the limited potentials that can be used with common solvents. Sensitivities in the 1 nM range have been demonstrated in small volumes (10 nL). The other advantage of this technique is that the number of electrons measured (through the current) is equal to the number of molecules present. The electrodes required for either of these detection methods can be included on a microfabricated device through a photolithographic patterning and metal deposition process. Electrodes could also be used to initiate a chemiluminescence detection process, i.e., an excited state molecule is generated via an oxidation-reduction process which then transfers its energy to an analyte molecule, subsequently emitting a photon that is detected.

[0092] Acoustic measurements can also be used for quantification of materials but have not been widely used to date. One method that has been used primarily for gas phase detection is the attenuation or phase shift of a surface acoustic wave (SAW). Adsorption of material to the

surface of a substrate where a SAW is propagating affects the propagation characteristics and allows a concentration determination. Selective sorbents on the surface of the SAW device are often used. Similar techniques may be useful in the methods described herein.

[0093] In certain embodiments, the methods of the present invention involve detection of laser activated fluorescence using microscope equipped with a camera. It is sometimes referred to as a high-efficiency photon detection system. Nie et. al. (1994), "Probing individual molecules with confocal fluorescence microscopy," *Science* 266:1018-1019.

[0094] The detection of single molecules involves limiting the detection to a field of view in which one has a statistical reason to believe there is only one molecule (homogeneous assays) or to a field of view in which there is only one actual point of attachment (heterogeneous assays). The single-molecule fluorescence detection of the present invention can be practiced using optical setups including near-field scanning microscopy, far-field confocal microscopy, wide-field epi-illumination, and total internal reflection fluorescence (TIRF) microscopy. For two-dimensional imaging fluorescence detection, the microscope is typically a total internal reflectance microscope. Vale et. al., 1996, Direct observation of single kinesin molecules moving along microtubules, *Nature* 380: 451, Xu and Yeung 1997, Direct Measurement of Single-Molecule Diffusion and Photodecomposition in Free Solution, *Science* 275: 1106-1109.

[0095] Suitable radiation detectors include may be, for example, an optical reader, photodiode, an intensified CCD camera, or a dye-impregnated polymeric coating on optical fiber sensor. In a preferred embodiment, an intensified charge couple device (ICCD) camera is used. The use of a ICCD camera to image individual fluorescent dye molecules in a fluid near the surface of the glass slide is advantageous for several reasons. With an ICCD optical setup, it is possible to acquire a sequence of images (movies) of fluorophores. In certain aspects, each of the NTPs of the present invention has a unique fluorophore associated with it, as such, a four-color instrument can be used having four cameras and four excitation lasers. Thus, it is possible to use this optical setup to sequence DNA. In addition, many different DNA molecules spread on a microscope slide can be imaged and sequenced simultaneously. Moreover, with the use of image analysis algorithms, it is possible to track the path of single dyes and distinguish them from fixed background fluorescence and from "accidentally dequenched" dyes moving into the field of view from an origin upstream.

[0096] In certain aspects, the preferred geometry for ICCD detection of single-molecules is total internal reflectance fluorescence (TIRF) microscopy. In TIRF, a laser beam totally reflects at a glass-water interface. The optical field does not end abruptly at the reflective interface, but its intensity falls off exponentially with distance. The thin "evanescent" optical field at the interface provides low background and enables the detection of single molecules with signal-to-noise ratios of 12:1 at visible wavelengths (*see*, M. Tokunaga *et al.*, *Biochem. and Biophys. Res. Comm.* **235**, 47 (1997) and P. Ambrose, *Cytometry*, **36**, 244 (1999)).

[0097] The penetration of the field beyond the glass depends on the wavelength and the laser beam angle of incidence. Deeper penetrance is obtained for longer wavelengths and for smaller angles to the surface normal within the limit of a critical angle. In typical assays, fluorophores are detected within about 200 nm from the surface which corresponds to the contour length of about 600 base pairs of DNA. Preferably, a prism-type TIRF geometry for single-molecule imaging as described by Xu and Yeung is used (*see*, X-H.N. Xu *et al.*, *Science*, **281**, 1650 (1998)).

[0098] Single molecule detection can be achieved using flow cytometry where flowing samples are passed through a focused laser with a spatial filter used to define a small volume. US Pat. No. 4,979,824 describes a device for this purpose. US Pat. No. 4,793,705 describes and claims in detail a detection system for identifying individual molecules in a flow train of the particles in a flow cell. The '705 patent further describes methods of arranging a plurality of lasers, filters and detectors for detecting different fluorescent nucleic acid base-specific labels. US Pat. No. 4,962,037 also describes a method for detecting an ordered train of labeled nucleotides for obtaining DNA and RNA sequences using a nuclease to cleave the bases rather than a polymerase to synthesize as described herein. Single molecule detection on solid supports is described in Ishikawa, *et al.* (1994) Single-molecule detection by laser-induced fluorescence technique with a position-sensitive photon-counting apparatus, *Jan. J. Apple. Phys.* **33**:1571-1576. Ishikawa describes a typical apparatus involving a photon-counting camera system attached to a fluorescence microscope. Lee *et al.* (1994), Laser-induced fluorescence detection of a single molecule in a capillary, *Anal. Chem.*, **66**:4142-4149 describes an apparatus for detecting single molecules in a quartz capillary tube. The selection of lasers is dependent on the label and the quality of light required. Diode, helium neon, argon ion, argon-krypton mixed ion, and Nd:YAG lasers are useful in this invention.

[0099] Detecting the fluorophore can be carried out using a variety of mechanisms. These mechanisms include for example, fluorescence resonance energy transfer, an electron transfer mechanism, an excited-state lifetime mechanism and a ground-state complex quenching mechanism.

5 F. **Labeled NTP residence times.**

[0100] The residence time of a correctly paired NTP (*i.e.*, an NTP that is complementary to the first unpaired nucleotide residue of the target NA that is just downstream from the extending end of the primer NA) is significantly longer than the residence time of an incorrectly paired NTP.

10 [0101] The kinetic mechanism has been well characterized for the reaction catalyzed by the T7 DNA polymerase. Patel *et al.* (1991), *Biochemistry* 30:511; Wong *et al.*, *Biochemistry* 30:526. In this reaction, the polymerase/target NA/primer NA complex is first contacted by an NTP. When a "correct" NTP (*i.e.*, complementary to the template nucleotide in the enzyme active site) binds, the enzyme pocket "closes" on the nucleotide and then the
15 coupling chemistry occurs. The enzyme "opens" back up, releases the PPi formerly attached to the NTP, and the enzyme translocates to the next base on the template. An incorrect NTP (*i.e.*, not complementary to the template base) has a very short residence time on the enzyme. See, *e.g.*, kinetic data at Table II of Patel *et al.* (1991), *Biochemistry* 30:511. In this instance and under the polymerization conditions used, the difference between an incorporated NTP
20 residence time is about 100 times longer to about 10,000 times longer than the residence time of an NTP that is not incorporated. In certain aspects, the residence time of an NTP that is incorporated into the primer nucleic acid is at least about 200 times longer to about 500 times longer such as 250, 350 or 450 times longer than the residence time of an NTP that is not incorporated.

25 [0102] The relatively long residence time of a correct NTP is used in the present invention to detect the interaction of a correct NTP with an immobilized polymerase/primer NA/template NA complex. Depending on the incubation conditions (*e.g.*, salt concentration, temperature, pH, *etc.*), the residence time of a nucleotide that is incorporated into an elongating primer is longer than the residence time of an NTP that is not incorporated. The
30 residence time of the label of a correct labeled NTP that is incorporated into the elongating primer ranges from about 1.0 milliseconds to about 100 milliseconds, preferably, from about 2.0 milliseconds to about 10 milliseconds. In certain instances, the accuracy of the residence time of the measurement depends on the speed of the detector.

III. EXAMPLES

Example 1. Introduce a unique cysteine on the protein surface for attaching a fluorophore

[0103] A unique cysteine amino acid is placed on the surface of Terminator polymerase to attach the fluorescent probe. This is accomplished by site-directed mutation of the Terminator gene in two steps. First, the single native surface-exposed cysteine, C223, is eliminated by mutation to serine, resulting in the mutant C223S. Mutant C223S has no surface-exposed cysteines. Next, a new cysteine is uniquely placed on the protein surface by constructing the mutant E554C. The new cysteine is located on the rim of a cleft in the protein, near the location of a quencher on a bound nucleotide. The resulting mutant is C223S:E554C.

Example 2. Add histidine patches to the protein surface attaching anchors

[0104] Two histidine patches are engineered onto the surface of the C223S:E554C Terminator protein by making the multiple mutations D50H:T55H:E189H:R196H:K229H. The resulting mutant, C223S:E554C:D50H:T55H:E189H:R196H:K229, is called "ThioHis".

Example 3. Circularization of target DNA

[0105] Randomly-sheared fragments of genomic DNA is purified from the sample organism. The DNA is treated with T4 DNA polymerase to generate blunt ends and a single "A" nucleotide is added to the 3'-ends with Taq DNA polymerase and dATP. A mixture of two double-stranded oligonucleotide adaptors is ligated to the DNA fragments with T4 DNA ligase. See, Figures 3-5.

First adaptor:

Biotin-CGCCACATTACACTTCCTAACACGT
GCGGTGTAATGTGAAGGATTGTGC

Second adaptor:

CAGTAGGTAGTCAAGGCTAGAGTCT
GTCATCCATCAGTTCGATCTCAG

Ligated DNA products:

genomic DNA: lower case
adaptors: upper case, (p) 5'-phosphate
italicized: DNA strand recovered after elution at alkaline pH

Product 1

Bio-CGCCACATTACACTTCCTAACACGTnnnnn...nnnnnaGACTCTAGCCTTGACTACCTACTGAAA-3'
GCGGTGTAATGTGAAGGATTGTGCannnnn...nnnnnTCGAGATCGGAAGTGGATGACp-5'

Product 2

Bio-CGCCACATTACACTTCCTAACACGTnnnnn...nnnnnaCGTGTTAGGAAGTGTAAATGTGGCG-3'
3'-GCGGTGTAATGTGAAGGATTGTGCannnnn...nnnnnTGACAATCCTTCACATTACACCGC-Bio

Product 3

5' -pCAGTAGGTAAGGCTAGAGTCTnnnnn...nnnnnGACTCTAGCCTTGACTACCTACTGAAA-3'
 3' -AAAGTCATCCATCAGTTCCGATCTCAGannnnn...nnnnnTCTGAGATCGGAACGTGATGGATGACp-5'

[0106] After ligation, DNA fragments in the size range of about 17-23 kb are purified by gel electrophoresis. The purified fragments are bound to streptavidin-coated magnetic beads (Dyna). After binding, the beads are washed to remove unbound DNA. Then the bound DNA is denatured at alkaline pH and the unbiotinylated strands are eluted (see above; Product 1, italicized font), and the DNA still bound to the beads is discarded. The eluted strands are circularized by hybridization to a primer oligo complementary to both adaptors:

Primed circular template

stars mark the ligation site: **
 5' -...nnnnnCGTGTAGGAAGTGTAATGTGGCGCAGTAGGTAGTCAAGGCTAGAGTCTnnnnn...-3' (template strand)
 3' -GCACAATCCTTCACATTACACCGCGTCATCCATCAGTTCCGATCTCAGA-5' (primer)

Example 4. Protein modifications

[0107] The ThioHis Terminator mutant protein (Example 2) is conjugated to tetramethylrhodamine-5-maleimide (Molecular Probes) at position C554. Anchors (biotin-X nitrilotriacetic acid, Molecular Probes) are added to bind to the two histidine patches and the modified protein is purified.

Example 5. Anchor protein-DNA complexes to glass coverslips

[0108] The modified ThioHis protein (Example 4) is mixed with the primed circular template DNA (Example 3) to form polymerase-DNA complexes. The complexes are added to a streptavidin-coated glass coverslip to topologically trap the DNA between the protein and the glass surface. The coverslip is washed prior to sequencing the immobilized DNA.

Example 6. Synthesis Of dUTP- γ -TMR**A. Synthesis of dUTP- γ S**

[0109] dUDP (16 mg, 40 μ mol; Sigma D-3626) and ATP-3S (44 mg, 80 μ mol; Boehringer Mannheim 102342) were dissolved in 10 mL of (20 mM Tris-Cl pH 7.0, 5% glycerol, 5 mM dithiothreitol, 5 mM MgCl₂). Nucleoside diphosphate kinase (0.5 mL, 5000 units; Sigma N-0379) was added and the sample was incubated at 37° C for 2 h to equilibrate the γ -thiophosphate moiety between the uridine and adenosine nucleotides. As expected from the reactant stoichiometry, 2/3 of the dUDP was converted to dUTP- γ S. The product was purified by reversed-phase HPLC using a linear gradient of 0% to 100% Buffer B mixed into

Buffer A (Buffer A is 0.1 M triethylammonium acetate in water, pH 7, 4% acetonitrile;
Buffer B is the same as Buffer A with 80% acetonitrile).

B. Synthesis of dUTP- γ -TMR

[0110] dUTP- γ S (45 μ g, 90 nmol; from step a) was dissolved in 295.5 μ L of (20 mM sodium phosphate pH 7.5, 33% dimethylformamide). BODIPY TMR1A (4.5 μ L, 0.45 μ mol dissolved in dimethylformamide; Molecular Probes) was added and the sample was held in the dark at room temperature for 2.5 h. The product was obtained in 90% yield and was purified by reversed-phase HPLC as in step a.

10 Example 7. Strep-Tag II T7 DNA Polymerase

[0111] The T7 DNA polymerase gene was amplified from T7 phage DNA using the forward primer

5'-ATGATCGTTTCTGCCATCGCAGCTAAC

(encodes the exonuclease mutations A14-to-C14 and A20-to-C20) and the reverse primer

15 5'-TCAGTGGCAAATCGCC.

[0112] An oligonucleotide encoding the Strep-Tag II sequence overlapping the 5'-end of the amplified T7 exo- polymerase gene was synthesized on an automated oligonucleotide synthesizer:

20 5'-ATGTCCAAC TGGTCCCACCCG CAGTTCGAAAAAGGTGGAGGTTCCGCT
M S N W S H P Q F E K G G G S A
Strep-Tag II Peptide Spacer

25 ATGATCGTTTCTGCCATCGCAGCTAAC..
M I V S A I A A N...
T7 polymerase N-terminus overlap (2 exo- mutations underlined)

[0113] The single-stranded synthetic oligonucleotide was spliced to the amplified T7 gene (above) by overlapping PCR (Horton *et al.* (1989) "Site-directed mutagenesis by overlap extension using the polymerase chain reaction," *Gene* 77:61-68) using the StrepTag forward primer

5'-ATGTCCAAC TGGTCCCACCC

with the reverse primer

5'-TCAGTGGCAAATCGCC.

[0114] The spliced PCR product was cloned into the pET11 plasmid vector (Stratagene), overexpressed in *E. coli* BL21(DE3)pLysS, and purified by Strep-Tag II affinity chromatography (Maier *et al.* (1998) *Anal. Biochem* 259: 68-73).

5 Example 8. Polymerase Immobilization

A. Surface passivation with polyethylene glycol

[0115] Fused silica coverslips (1" square, 200 μ m thick; SPI Supplies, West Chester PA) were cleaned by soaking overnight in chromic acid and washing in distilled water in a sonic bath (Model 2200, Branson, Danbury CT). Methoxy-PEG-silane MW 5,000 (Shearwater
10 Polymers, Huntsville AL) was dissolved at 10 mg/ml in 95:5 ethanol:water and the pH was adjusted to 2.0 with HCl. Cleaned coverslips were immersed in the PEG solution for 2 hours, washed 3 times each in ethanol, 3 times in water, dried overnight at 70 C, washed overnight in 1% sodium dodecyl sulfate in water, washed with deionized water in an ultrasonic bath, and baked for 1 day at 70 C (Jo S, Park K. Surface modification using silanated
15 poly(ethyleneglycol)s. *Biomaterials* 21: 605-616. 2000).

B. Biotinylation and streptavidin monolayer

[0116] Photoactivatable biotin (12 μ g; Pierce, Rockford IL) was dissolved in 1 ml of deionized water. The solution was applied to the top surface of a PEG-silane coated coverslip from step (a) and the water was evaporated under vacuum. The coverslip was
20 exposed to UV light (General Electric Sunlamp RSM, 275W) for 20 minutes at a distance of 5 cm. The coverslip was washed with deionized water and nonspecific binding sites are blocked by overlaying a solution of 3% bovine serum albumin in 50 mM Tris-Cl pH 7.5, 150 mM NaCl (TBS) for 1 hour at room temperature. The coverslip was washed with TBS, a solution of streptavidin (1 mg / mL in TBS; Pierce, Rockford IL) was applied for 30 minutes,
25 and the coverslip was washed with TBS + 0.1% Tween 20 followed by TBS alone.

[0117] The streptavidin-coated coverslip from step (b) was spotted with 20 μ L of T7 DNA polymerase exo⁻ Strep-tag II (10 μ M in TBS). After 1 hr, the coverslip was washed with TBS, ready for use.

C. Nickel nanodots

30 [0118] In one embodiment, a polymerase is attached to each dot of an array of nickel nanodots. (Depending on the fluorophore used, the nickel nanodot may, however, exhibit background fluorescence, which must be corrected for.) The required equipment includes a spinner (PWM 202 E-beam resist spinner, Headway Research Inc.), an evaporator (SC4500

thermal e-gun evaporator, CVC Products Inc.), and a scanning electron microscope (Leo 982 with Nabity pattern generator, Leo Electron Microscopy Inc.).

[0119] Clean a 25 mm diameter microscope coverslip on the spinner by spraying alternately with acetone and isopropyl alcohol (IPA) and spinning the last IPA film until dry.

5 Coat the coverslip in the spinner with 0.5 ml of PMMA (poly(methyl methacrylate), MW 496 kDa, 2% in chlorobenzene), bake on a hotplate at 170 C for 10 min, coat with 0.5 ml of PMMA (MW 950 kDa, 2% in methyl isobutyl ketone [MIBK]), and bake again. Apply the conductive layer by evaporating 100 Angstroms of gold onto the PMMA film in the CVC SC4500. Use the electron microscope to etch the array pattern into the PMMA film using a
10 pattern generator on the Leo 982 as specified by a CAD drawing (Design CAD, 50 nm spots, 10 μ m center-to-center spacing, 200 x 200 dot array).

[0120] Remove the gold layer by placing the exposed coverslip in Gold Etch (15-17% sodium iodide) for 7 seconds followed by rinsing with IPA and water. Deposit Tantalum (50 Angstroms) and Nickel (100 Angstroms) on the coverslip in the CVC SC4500. Remove the
15 PMMA in a 1:1 mix of acetone and methylene chloride for 10-15 min followed by sonication for several seconds and rinsing with IPA and water.

[0121] Attach the polymerase just before use by applying 10 μ l of a 15 nM solution of polyhistidine-tagged Klenow DNA polymerase exo^- (prepared using TOPO cloning vector and ProBond Resin, Invitrogen Inc.) in phosphate-buffered saline (PBS; Harlow E., Lane D.
20 1988. Antibodies A Laboratory Manual. Cold Spring Harbor Laboratory ISBN 0-87969-14-2) to the coverslip; after 20 min, wash the coverslip in PBS and use immediately.

Example 9. Determination Of Cystic Fibrosis Mutant

[0122] A polymerase-coated coverslip is placed on the microscope and a 20 μ l sample is
25 applied under a water immersion objective lens. The sample contains 40 mM Tris-Cl (pH 7.5), 1 mM ethylenediaminetetraacetic acid, 1 mM dithiothreitol, 0.1 mg/ml of bovine serum albumin, 12.5 mM magnesium chloride, 10 nM dUTP-TMR, 100 nM each of dATP, dCTP, and dGTP, and 10 μ g/ml of primer-template DNA. Depending on the activity of the immobilized enzymes, the nucleotide concentration may have to be adjusted so that
30 individual incorporation events are time-resolvable. Data are collected and analyzed as described in Example 6 to determine whether the dUTP-TMR nucleotide is incorporated into the primer strand. (In order to perform this experiment in a droplet on an open coverslip as described, it may be necessary to speed the motion of free dUTP-TMR through the imaged zone by drive convection with a nitrogen stream, depending on ambient conditions. It is also

necessary to use a water immersion objective lens immersed directly in the sample.) The results are compared against a control without primer-template DNA to demonstrate the appearance of longer fluorescence bursts in the test sample indicating a template sequence which supports dUTP incorporation. Two sample primer-templates are compared; they are synthetic oligonucleotides derived from the cystic fibrosis transmembrane conductance regulator gene (Welsh *et al.* (1993), *J. Cell Science* 106S:235-239).

Normal Allele (does not incorporate dUTP- γ -TMR)

primer 3'-CACCATTAAGAAAATATCAT

template 5'-GUGGUAAUUUCUUUUAUAGUAG

(DELTA)F508 DELETION MUTANT (DOES INCORPORATE DUTP- γ -TMR)

primer 3'-CACCATTAAGAAAATATCAT

template 5'-GUGGUAAUUUCUUUUAUAGUAA

Example 10. Microscope Setup

[0123] The setup for a residence-time detector is described in FIG. 10. A multicolor mixed-gas laser 1 emits light at tunable wavelengths. The laser beam is first passed through a laser line filter 2 and then at a right angle into a fused-silica prism 3 which is optically connected to the fused silica flowcell 4 by immersion oil. The labeled nucleotides 6 flow in a buffer solution across the polymerase enzymes immobilized on the surface of the flowcell chamber 7. Laser light strikes the fused silica-buffer interface at an angle such that the critical angle between fused-silica and the buffer solution is exceeded. The light is thus completely reflected at the interface, giving rise to a total internal reflection (TIR) evanescent field 5 in the solution. The angle is adjusted to give a 1/e penetrance of between 1 and 200 nm into the solution. The immobilized polymerases 7 are illuminated in the evanescent field and are imaged using a microscope 9 with an objective lens 8 mounted over the flowcell. Fluorescence emission at the microscope output passes through a notch filter 10 and a long pass filter 11 which allow the fluorescence emission to pass through while blocking scattered laser light. The fluorescence photons are focused onto a single-photon avalanche diode SPAD 12. Signals are processed by a constant fraction discriminator CFD 13, digitized by an analog-to-digital converter ADC 14, and stored in memory 15. Signal extraction algorithms 16 are performed on the data stored in memory. These algorithms may distinguish signal from background, filter the data, and perform other signal processing functions. The signal processing may be performed off-line in a computer, or in specialized digital signal processing (DSP) chips controlled by a microprocessor. The fluorescence is recorded using,

for example by using CCD camera capable of recording single fluorophore molecules. Residence times and polymerase speed may be manipulated by controlling the reaction conditions (temperature, pH, salt concentration, labeled NTP concentration, etc.)

5 Example 11. Data Acquisition And Analysis

[0124] A computer model was developed to show the appearance of known (*i.e.*, simulated) incorporation events where the nucleotide is retained by a polymerase while the base-addition chemistry occurs.

[0125] The simulation was written in MATLAB. It operates by introducing free
10 background nucleotides into the field of view at a rate determined by the flux, which is calculated from the bias flow and optical detection volume. The detection volume is determined by the diffraction-limited focus (Airy disc diameter) and depth of the evanescent light field. The time between molecule arrivals is governed by an exponential probability distribution. As each molecule enters the simulation, the number of photons it emits is a
15 Poisson random number, with mean calculated from the time it spends in the focal volume (determined by the bias flow), the excitation rate of the molecule (determined by the laser intensity, photon energy, and absorption cross section of the dye), and the fluorescence quantum yield of the dye. The number of photons seen by the detector is calculated in turn by the detection efficiency ratio. The photons detected are scattered in time according to a
20 second exponential distribution, with rate calculated from the photon capture rate.

[0126] Signal molecules (*i.e.*, nucleotides bound to the enzyme during the base-addition reaction) are introduced in time at a rate given by another simulation parameter, the reaction rate, and again distributed by a separate exponential distribution. The time a signal molecule spends in the resolution volume is determined by a random number with uniform distribution
25 from 2 to 5 ms, consistent with the enzyme kinetics of T7 DNA polymerase (Patel S, Wong I, Johnson K (1991) Biochemistry 30: 511). The number of photons detected is a Poisson random number with mean detected as in the background molecule case. The photons detected are distributed according to the same distribution as the photons coming from background molecules.

30 [0127] To detect the residence-time bursts, the time arrival of all photons is discretized by a sample clock. Then the photon data is processed with a weighted sliding-sum filter, using a Hamming window. The signal energy is calculated and displayed in time. The bursts are detected by two thresholds: a signal energy threshold (vertical), and a time threshold

(horizontal). A photon burst must pass both thresholds in order to be classified as a signal event.

[0128] Two simulation results are shown in Figures 11 and 12. The parameters are the same between the two Figures (Table III).

5

Table III

PARAMETER NAME	VALUE
Laser power	150 (mW)
Laser spot diameter	20 (micrometers)
Numerical aperture of objective lens	1.2
Evanescence light field height	30 (nm)
Bias flow	2 (mm/s)
Molarity	10e-9 (mol/L)
Fluorescence quantum yield (for Tetramethylrhodamine, TMR)	0.15
Net detection efficiency	3%
Sample clock	1.0 (MHz)

[0129] As is shown in FIG. 11, six incorporation events have occurred, all of the
10 incorporation events are detected above a signal energy threshold of 2500. FIG. 12 corresponds to photon data from background molecules only. Figures 11 and 12 clearly illustrate that incorporation events and the identity of incorporated NTPs can be detected by measuring NTP residence times.

[0130] All publications and patent applications cited in this specification are herein
15 incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

WHAT IS CLAIMED IS:

1 1. A polymerase-nucleic acid complex, said polymerase-nucleic acid
2 complex comprising:
3 a target nucleic acid and a nucleic acid polymerase, wherein said polymerase
4 has an attachment complex comprising at least one anchor, which said at least one anchor
5 irreversibly associates said target nucleic acid with said polymerase to increase the
6 processivity index.

1 2. The polymerase-nucleic complex of claim 1, wherein said polymerase-
2 nucleic acid complex further comprises a primer nucleic acid which complements a region of
3 said target nucleic acid.

1 3. The polymerase-nucleic complex of claim 1, wherein said attachment
2 complex comprises at least two anchors.

1 4. The polymerase-nucleic complex of claim 3, wherein said attachment
2 complex is attached to a support.

1 5. The polymerase-nucleic complex of claim 1, wherein said attachment
2 complex comprises a topological tether.

1 6. The polymerase-nucleic complex of claim 3, wherein said at least two
2 anchors further comprises a topological tether.

1 7. The polymerase-nucleic complex of claim 6, wherein said topological
2 tether is attached to at least one anchor via a complementary binding pair.

1 8. The polymerase-nucleic complex of claim 6, wherein said topological
2 tether is attached to at least two anchors via at least two complementary binding pairs.

1 9. The polymerase-nucleic complex of claim 7, wherein said
2 complementary binding pairs are selected from the group consisting of any haptenic or
3 antigenic compound in combination with a corresponding antibody or binding portion or
4 fragment thereof, nonimmunological binding pairs, receptor-receptor agonist or antagonist,
5 IgG-protein A, lectin-carbohydrate, enzyme-enzyme cofactor, enzyme-enzyme-inhibitor, and
6 complementary polynucleotide pairs capable of forming nucleic acid duplexes.

1 10. The polymerase-nucleic complex of claim 9, wherein said
2 complementary binding pair is selected from the group consisting of digoxigenin and anti-
3 digoxigenin, fluorescein and anti-fluorescein, dinitrophenol and anti-dinitrophenol,
4 bromodeoxyuridine and anti-bromodeoxyuridine, mouse immunoglobulin and goat anti-
5 mouse immunoglobulin, biotin-avidin, biotin-streptavidin, thyroxine and cortisol, a
6 phenylalanine derivative and hydrazine linker and acetylcholine and receptor-acetylcholine.

1 11. The polymerase-nucleic complex of claim 1, wherein said at least one
2 anchor comprises at least one amino acid or an epitope for attachment.

1 12. The polymerase-nucleic complex of claim 11, wherein said at least one
2 amino acid is selected from the group consisting of a cysteine, a phenylalanine derivative and
3 a histidine.

1 13. The polymerase-nucleic complex of claim 12, wherein said histidine is
2 selected from the group consisting of a histidine tag, a histidine patch and a polyhistidine
3 sequence.

1 14. The polymerase-nucleic complex of claim 5, wherein said topological
2 tether comprises an antibody.

1 15. The polymerase-nucleic complex of claim 1, wherein said at least one
2 anchor is attached to a support.

1 16. The polymerase-nucleic complex of claim 1, wherein said at least one
2 anchor entraps said target nucleic acid.

1 17. The polymerase-nucleic complex of claim 6, wherein said topological
2 tether is an antibody and said at least two anchors are each a histidine tag.

1 18. The polymerase-nucleic complex of claim 1, wherein said target
2 nucleic acid is a circular DNA.

1 19. The polymerase-nucleic complex of claim 18, wherein said circular
2 DNA is sequenced by strand displacement synthesis.

1 20. The polymerase-nucleic complex of claim 1, wherein said polymerase
2 is a selected from a Family A polymerase and a Family B polymerase.

1 21. The polymerase-nucleic complex of claim 20, wherein said Family A
2 polymerase is selected from the group consisting of Klenow, Taq, and T7 polymerase.

1 22. The polymerase-nucleic complex of claim 20, wherein said Family B
2 polymerase is selected from the group consisting of a terminator polymerase, phi29, RB-69
3 and T4 polymerase.

1 23. The polymerase-nucleic complex of claim 1, wherein said polymerase-
2 nucleic acid complex is an array of polymerase-nucleic acid complexes attached to a support.

1 24. The polymerase-nucleic complex of claim 23, wherein the plurality of
2 members of said array of polymerase-nucleic acid complexes is randomly attached to said
3 support.

1 25. The polymerase-nucleic complex of claim 23, wherein the plurality of
2 members of said array of polymerase-nucleic acid complexes is uniformly attached to said
3 support.

1 26. The polymerase-nucleic complex of claim 1, wherein the processivity
2 index is at least 0.5.

1 27. The polymerase-nucleic complex of claim 26, wherein the processivity
2 index is at least 0.8.

1 28. The polymerase-nucleic complex of claim 27, wherein the processivity
2 index is 1.

1 29. A method for detecting incorporation of at least one NTP into a single
2 primer nucleic acid molecule, said method comprising:

3 i. immobilizing onto a support a polymerase nucleic acid complex
4 comprising a target nucleic acid, a primer nucleic acid which complements a region of the
5 target nucleic acid, and at least one nucleic acid polymerase;

6 ii. contacting said immobilized complex with at least one type of labeled
7 nucleotide triphosphate [NTP], wherein each NTP is labeled with a detectable label, and

8 iii. detecting the incorporation of said at least one type of labeled NTP into
9 a single molecule of said primer, while said at least one type of labeled NTP is in contact
10 with said immobilized complex, by detecting the label of the NTP while said at least one type
11 of labeled NTP is in contact with said polymerase nucleic acid complex.

1 30. The method of claim 29, wherein said polymerase nucleic acid
2 complex is contacted with a single type of labeled NTP.

1 31. The method of claim 29, wherein said polymerase nucleic acid
2 complex is contacted with at least two different types of NTPs, and wherein each type of
3 NTP is uniquely labeled.

1 32. The method of claim 29, wherein said polymerase nucleic acid
2 complex is contacted with at least four different types of NTPs, and wherein each type of
3 NTP is uniquely labeled.

1 33. The method of claim 29, wherein said NTPs are labeled on the γ -
2 phosphate.

1 34. The method of claim 33, wherein said NTPs are labeled on the γ -
2 phosphate with a fluorescent label.

1 35. The method of claim 29, wherein the detecting comprises detecting a
2 unique signal from the labeled NTP using a system or device selected from the group
3 consisting of an optical reader, a high-efficiency photon detection system, a photodiode, a
4 camera, a charge couple device, an intensified charge couple device, a near-field scanning
5 microscope, a far-field confocal microscope, a microscope that detects wide-field epi-
6 illumination, evanescent wave excitation and a total internal reflection fluorescence
7 microscope.

1 36. The method of claim 29, wherein the label of the NTP is detected using
2 a method comprising a four color evanescent wave excitation device.

1 37. The method of claim 29, wherein said detecting is carried out by a
2 mechanism selected from the group consisting of fluorescence resonance energy transfer, an
3 electron transfer mechanism, an excited-state lifetime mechanism and a ground-state complex
4 quenching mechanism.

1 38. The method of claim 29, wherein said detecting step comprises
2 measuring a residence time of a labeled NTP in said polymerase nucleic acid complex.

1 39. The method of claim 38, wherein the residence time of an NTP that is
2 incorporated into the primer nucleic acid is at least about 100 times longer to about 10,000
3 times longer than the residence time of an NTP that is not incorporated.

1 40. The method of claim 39, wherein the residence time of an NTP that is
2 incorporated into the primer nucleic acid is at least about 200 times longer to about 500 times
3 longer than the residence time of an NTP that is not incorporated.

1 41. The method of claim 38, wherein the residence time of an NTP that is
2 incorporated into the primer nucleic acid is about 1.0 milliseconds to about 100 milliseconds.

1 42. The method of claim 41, wherein the residence time of an NTP that is
2 incorporated into the primer nucleic acid is about 2.0 milliseconds to about 10 milliseconds.

1 43. The method of claim 29, further comprising the step of
2 genotyping said target nucleic acid by determining the identity of at least one
3 NTP that is incorporated into a single molecule of the primer.

1 44. The method of claim 29, further comprising: sequencing said target
2 nucleic acid by determining the identity and sequence of incorporation of NTPs that are
3 incorporated into a single molecule of the primer.

1 45. The method of claim 29, wherein said detection is a sequential
2 detection of the identities of more than one uniquely labeled dNTPs that are sequentially
3 incorporated into the primer, wherein said sequential detection yields the sequence of region
4 of the target DNA that is downstream of the elongating end of the primer.

1 46. The method of claim 29, wherein said polymerase-nucleic acid
2 complex comprises a target nucleic acid and a nucleic acid polymerase, wherein said
3 polymerase has an attachment complex comprising at least one anchor, which irreversibly
4 associates said target nucleic acid with said polymerase for increasing the processivity index.

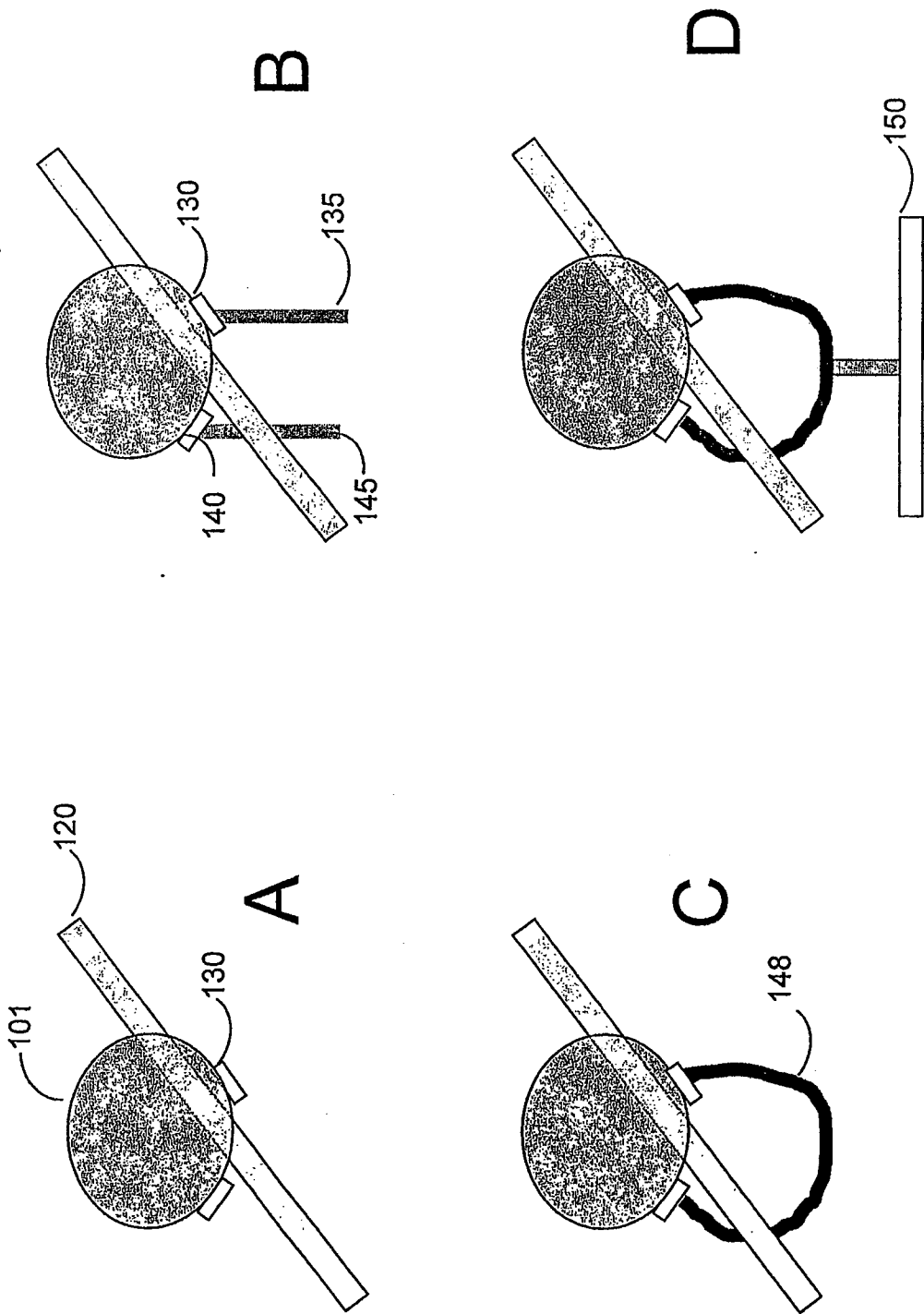


FIG. 1

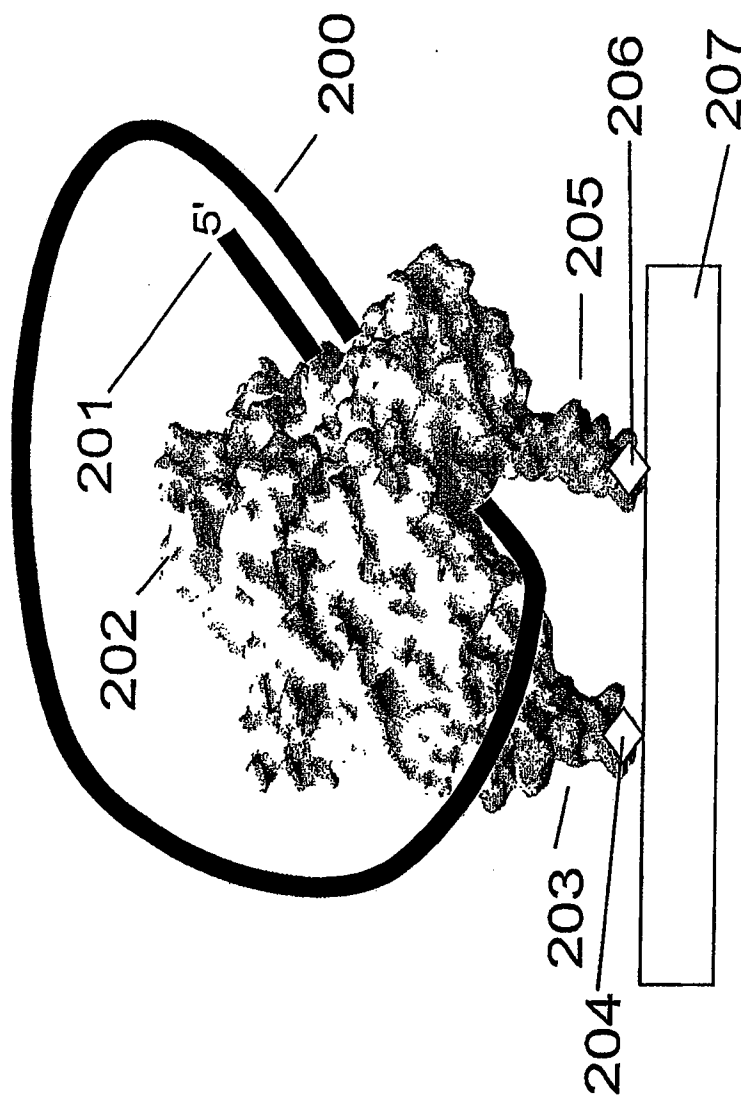
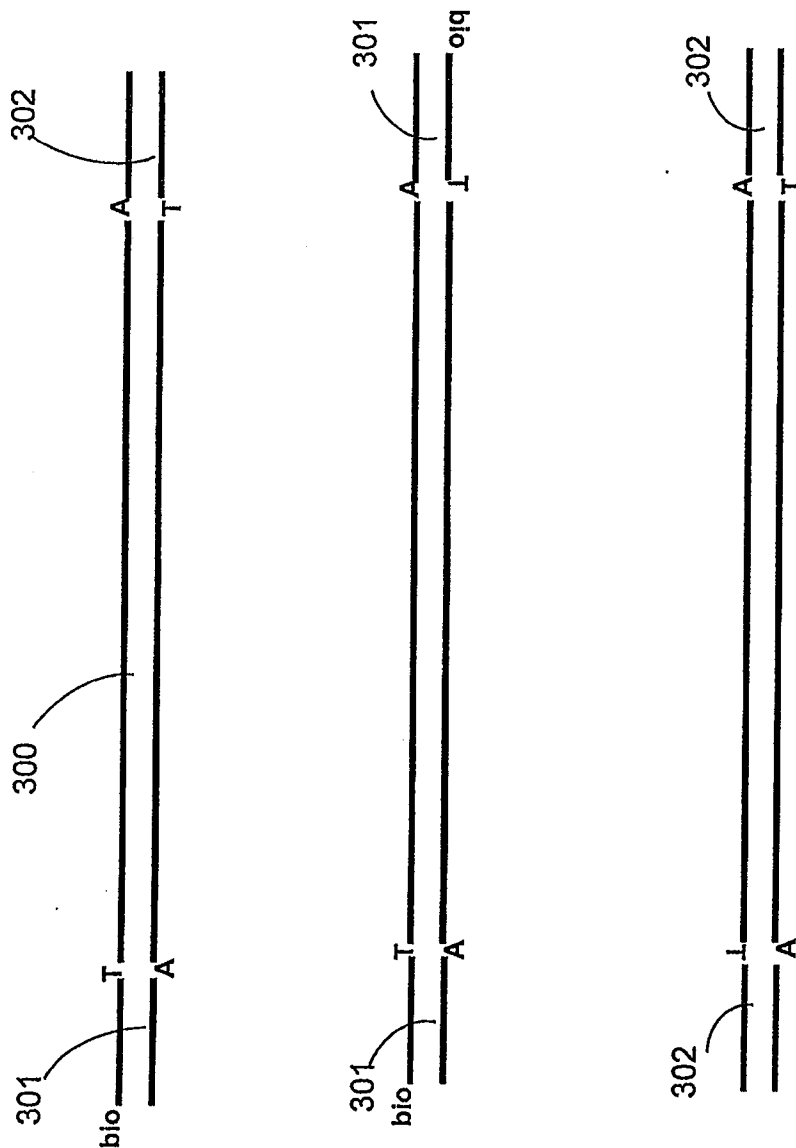


FIG. 2



DNA Sample Prep

adaptor ligation



PURIFY 20KB FRAGMENTS

FIG. 3

DNA Sample Prep

strand isolation

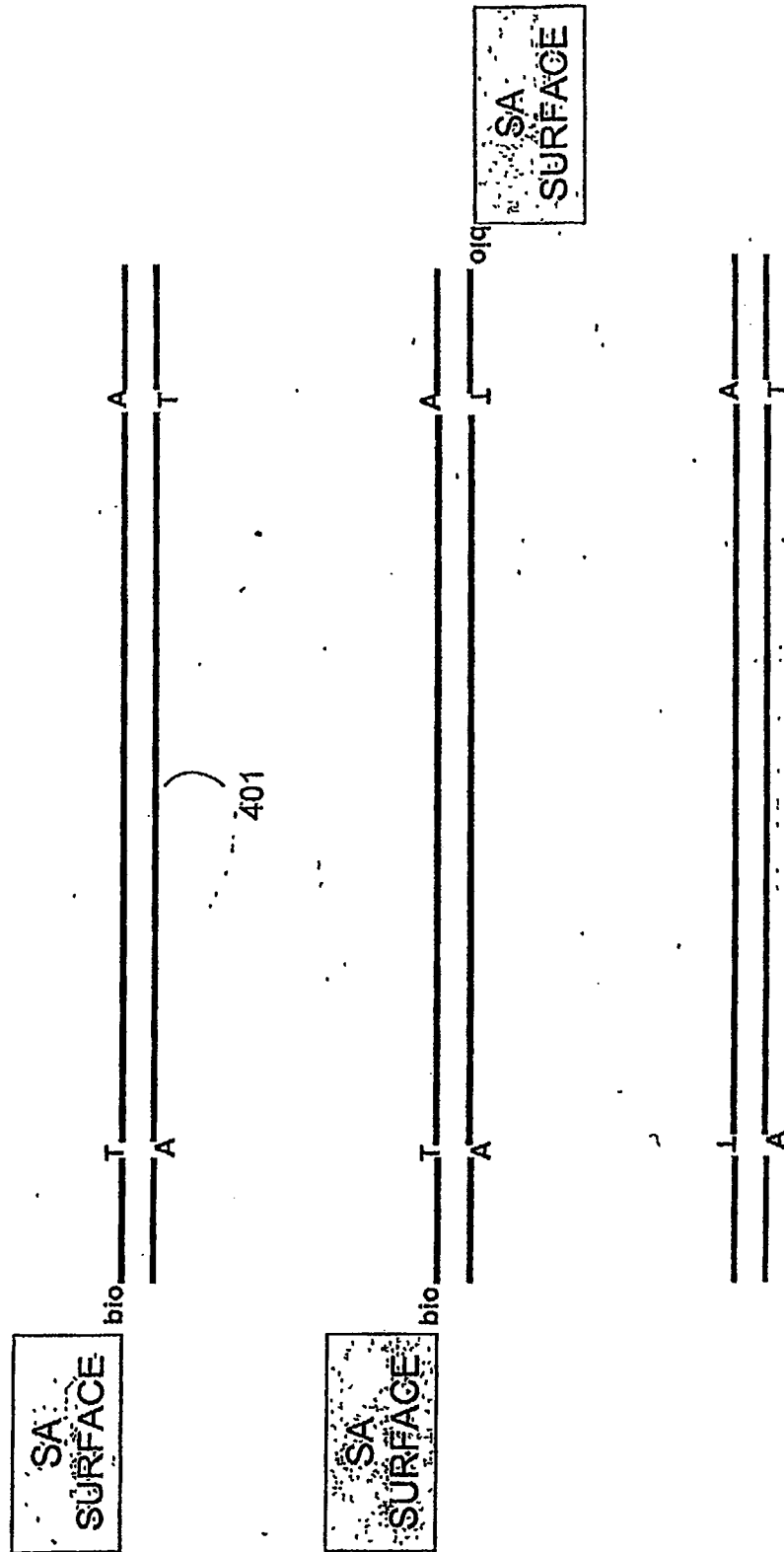


FIG. 4

DNA Sample Prep

circularization

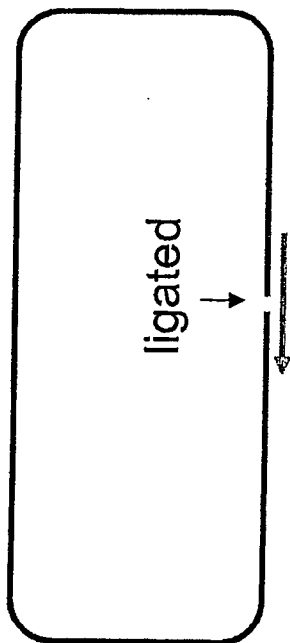


FIG. 5

Single Molecule Configuration DNA-protein complexes

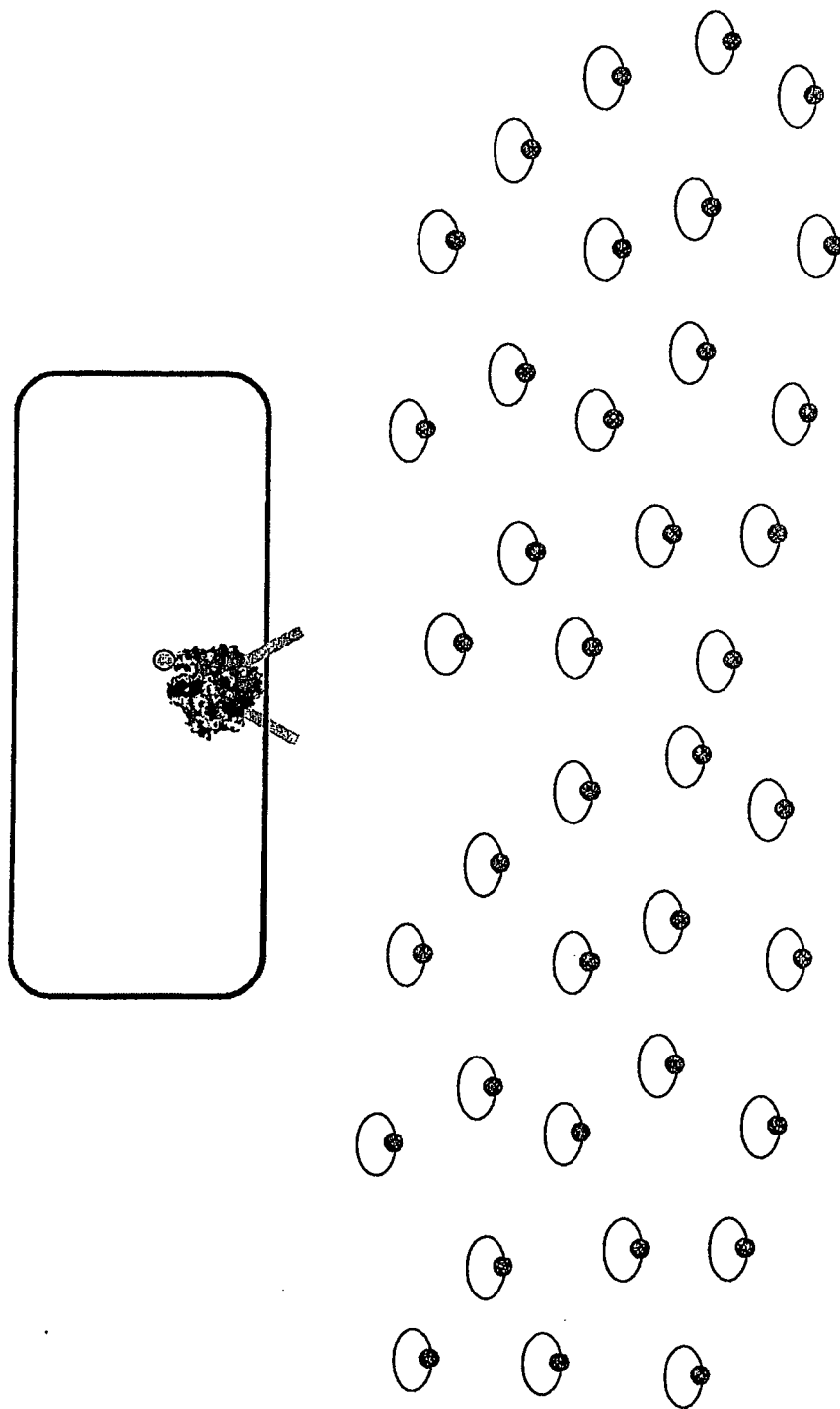


FIG. 6

Single-Molecule Isolation

bind to coverslip

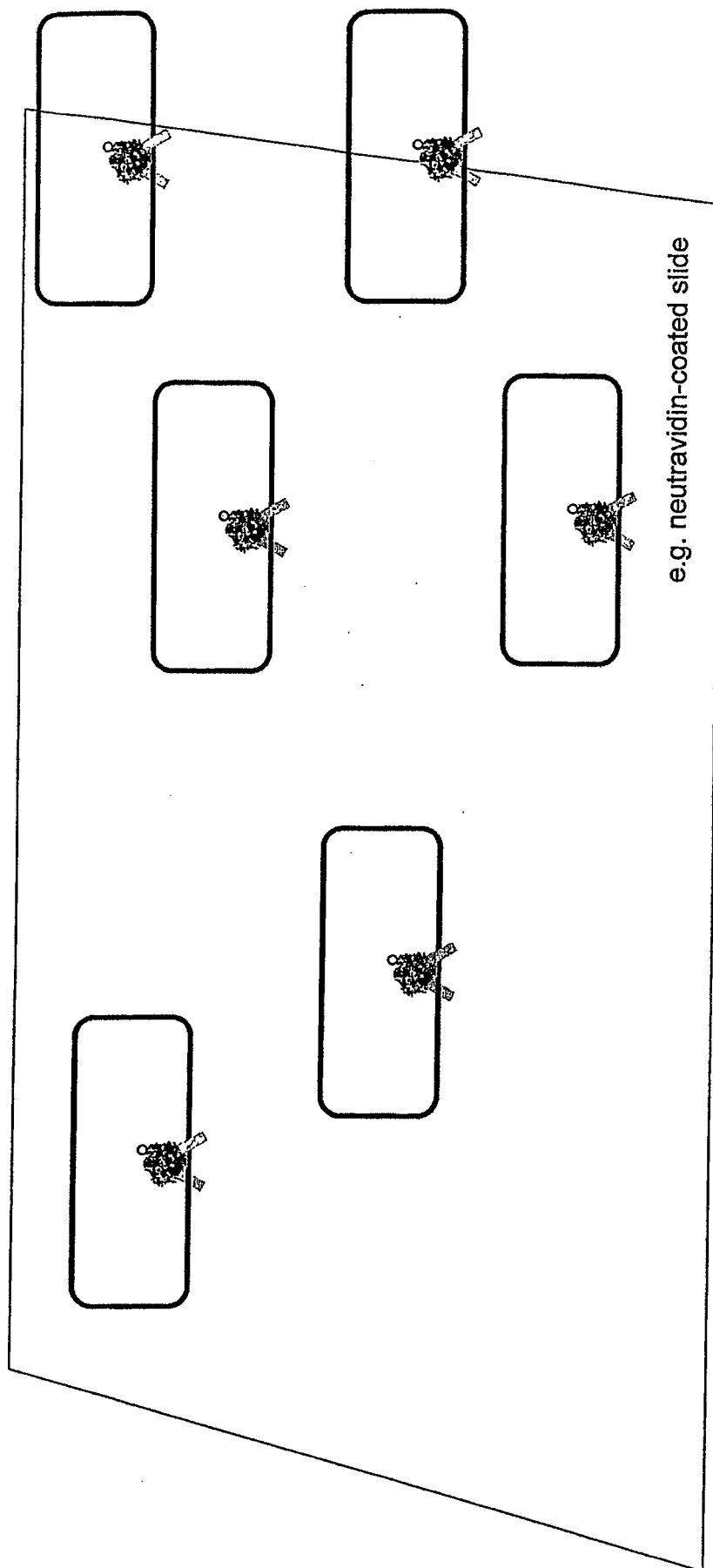


FIG. 7

Sequence hundreds of individual single molecules in each well of a microtiter plate

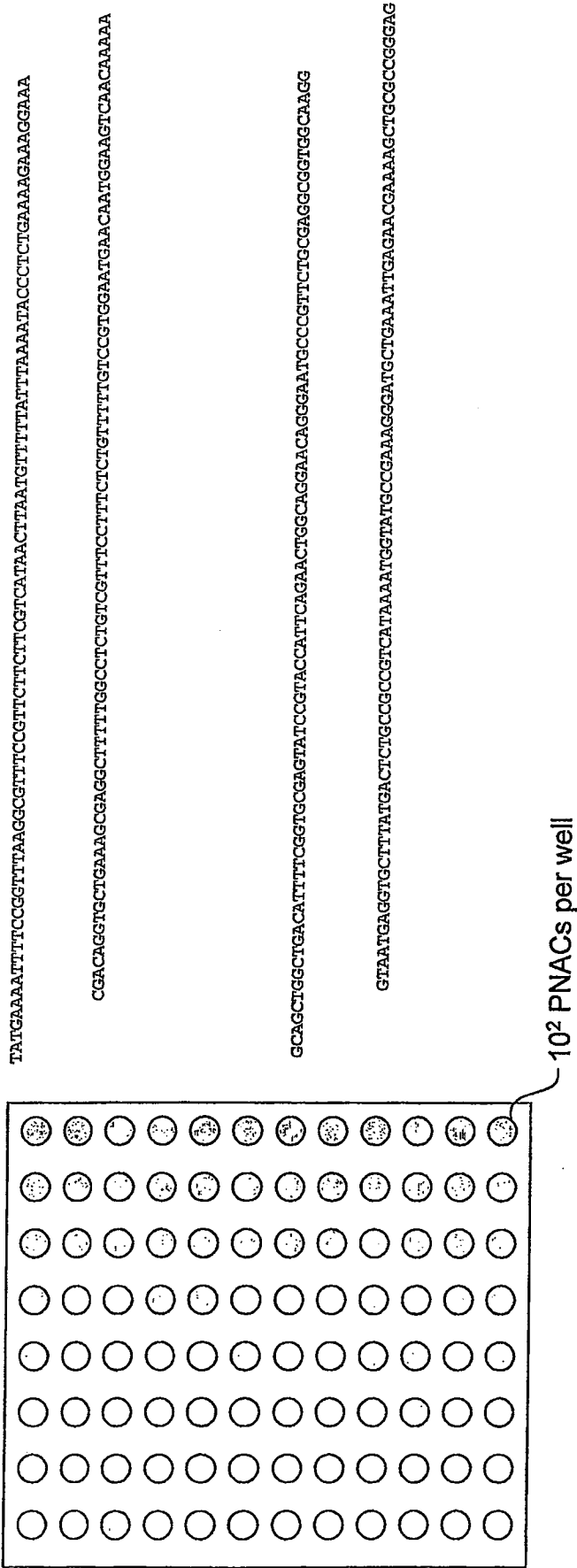


FIG. 8

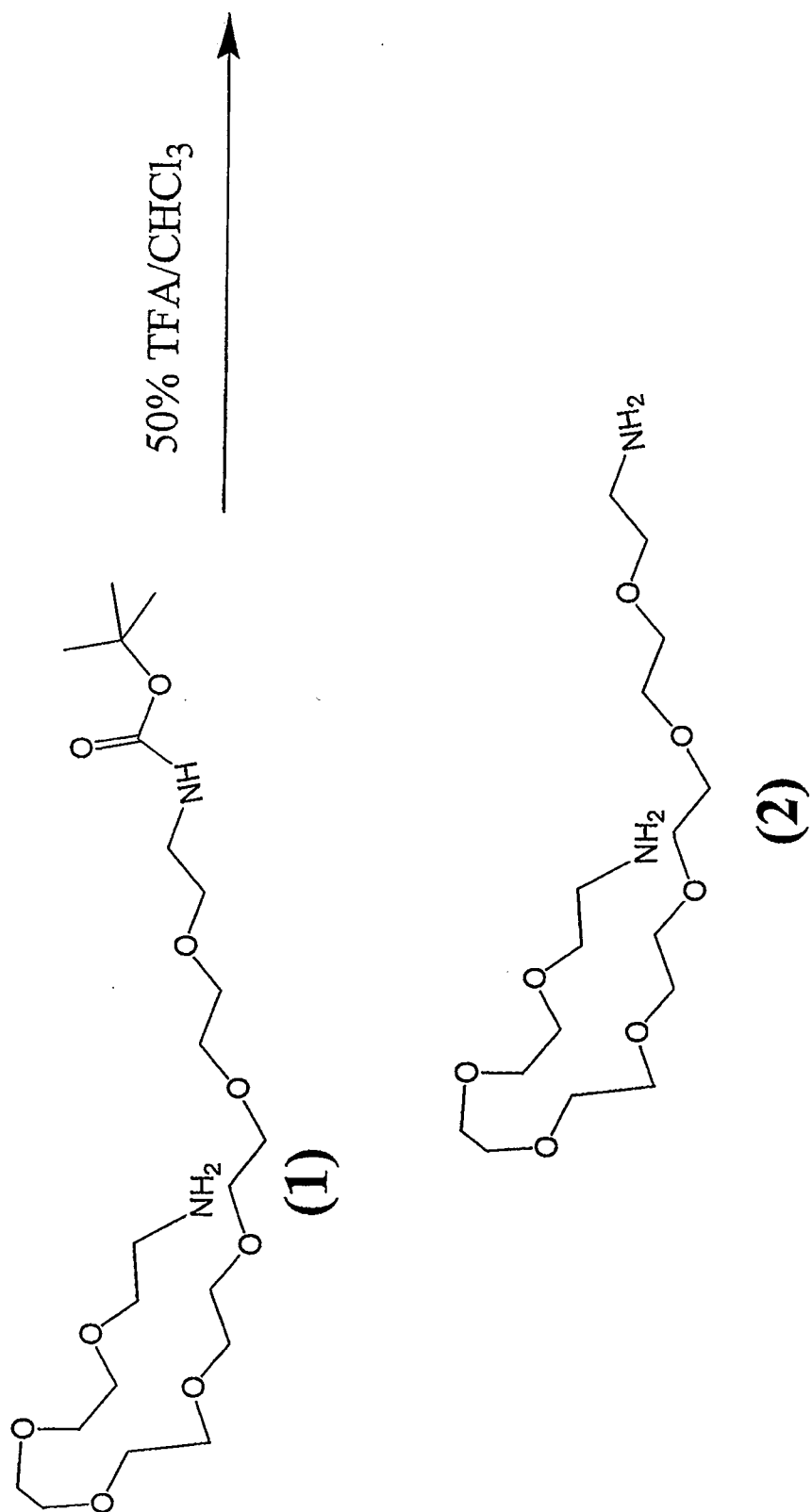


FIG. 9A

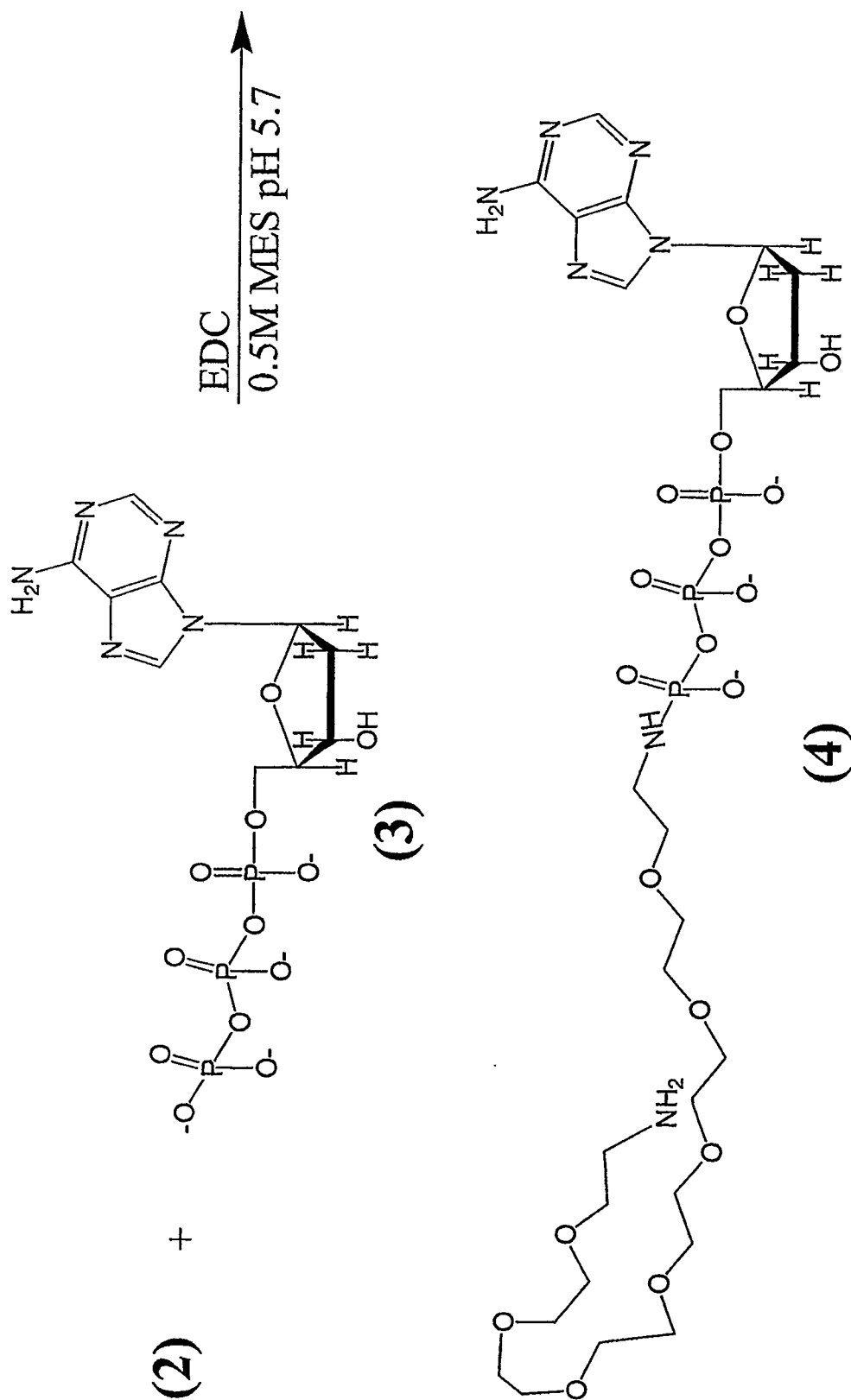


FIG. 9B

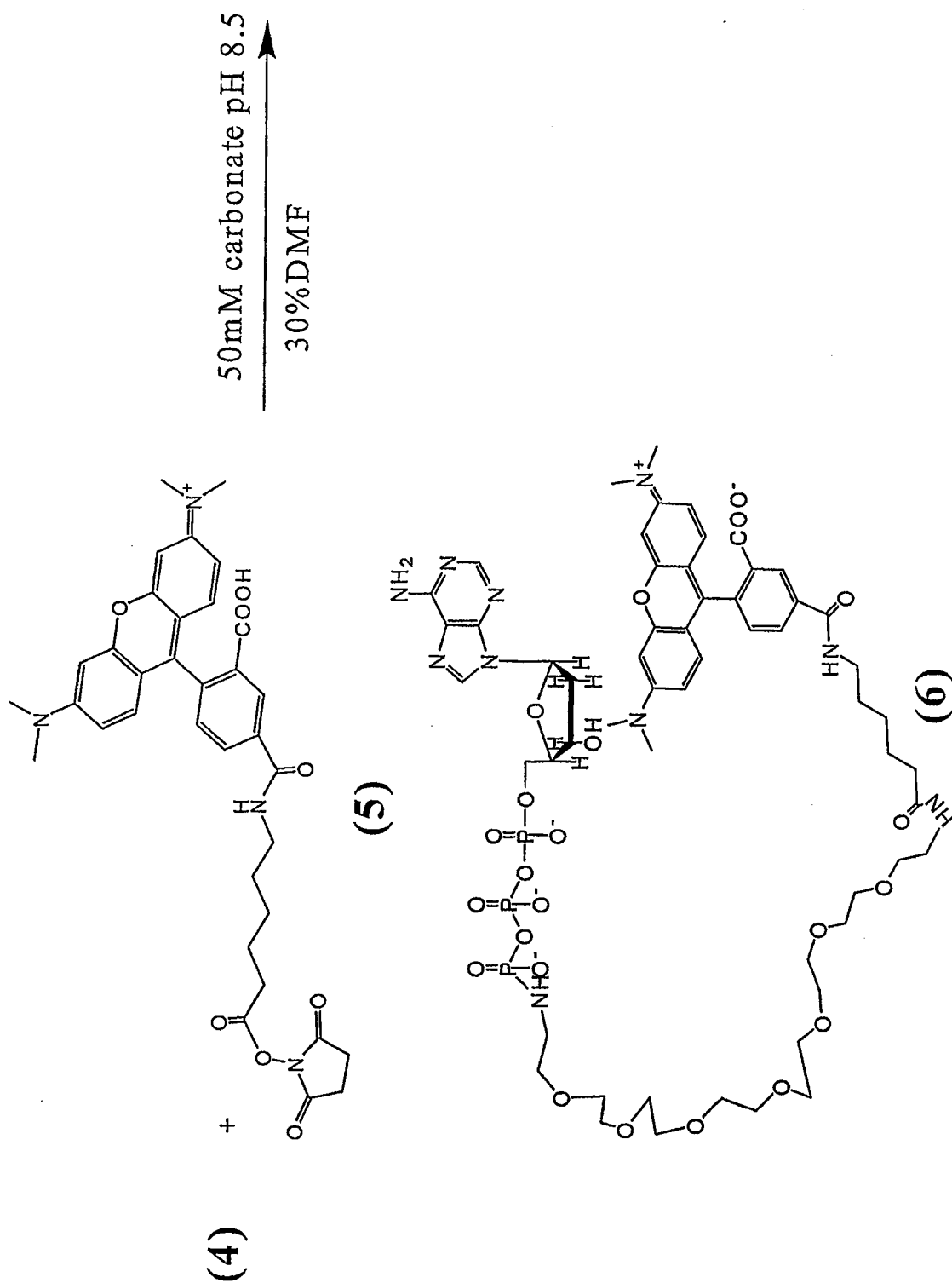


FIG. 9C

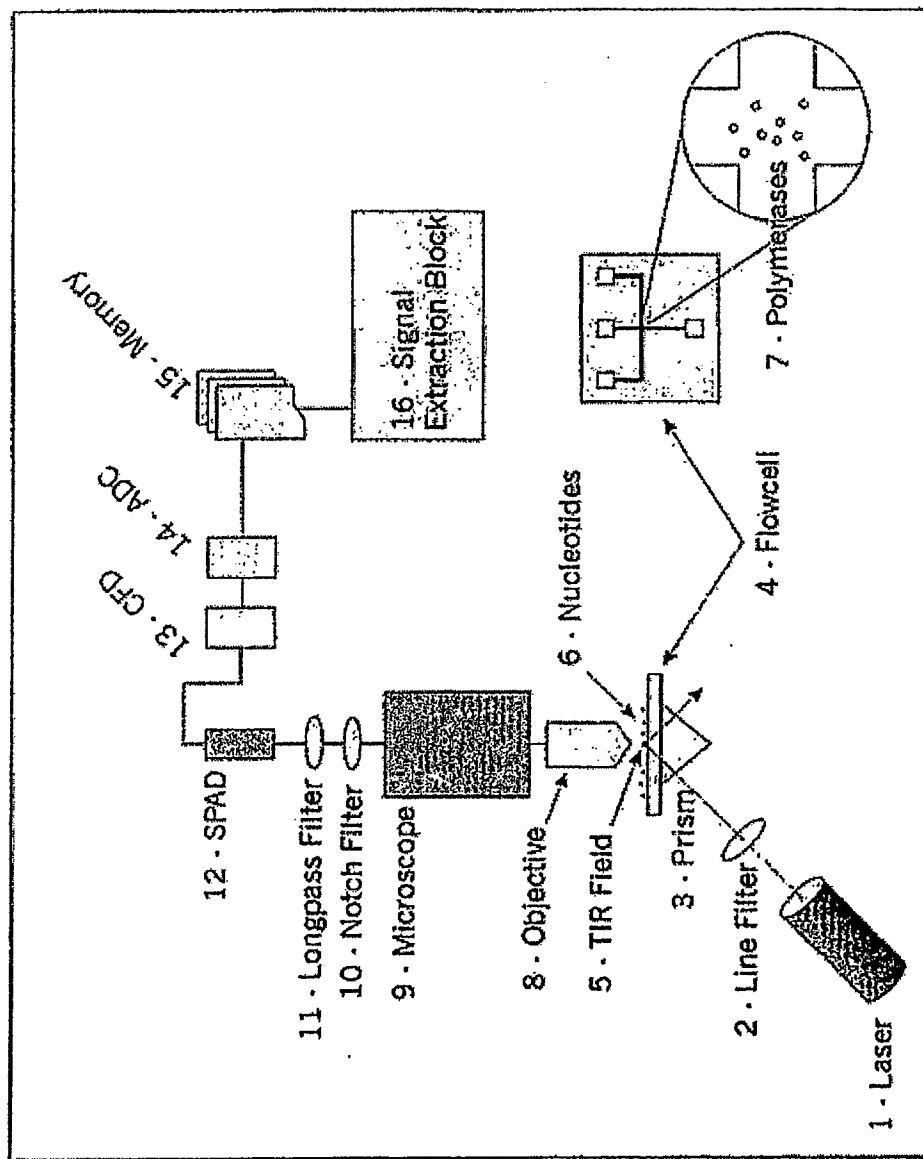


FIG. 10

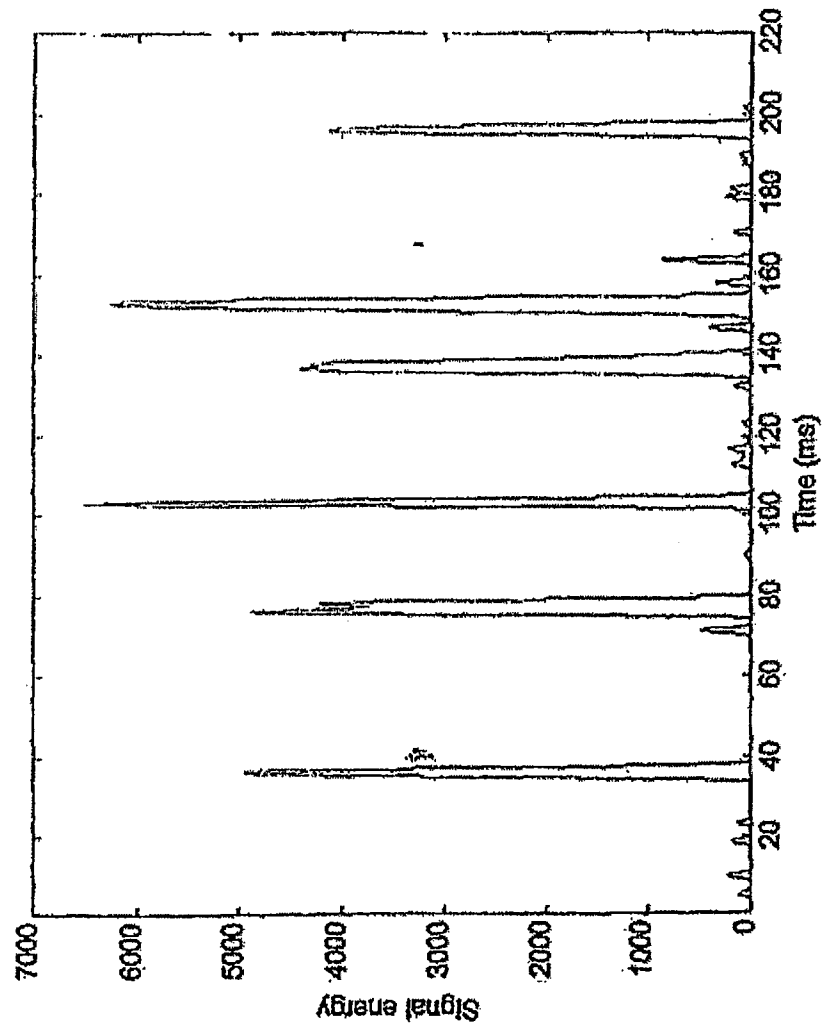


FIG. 11

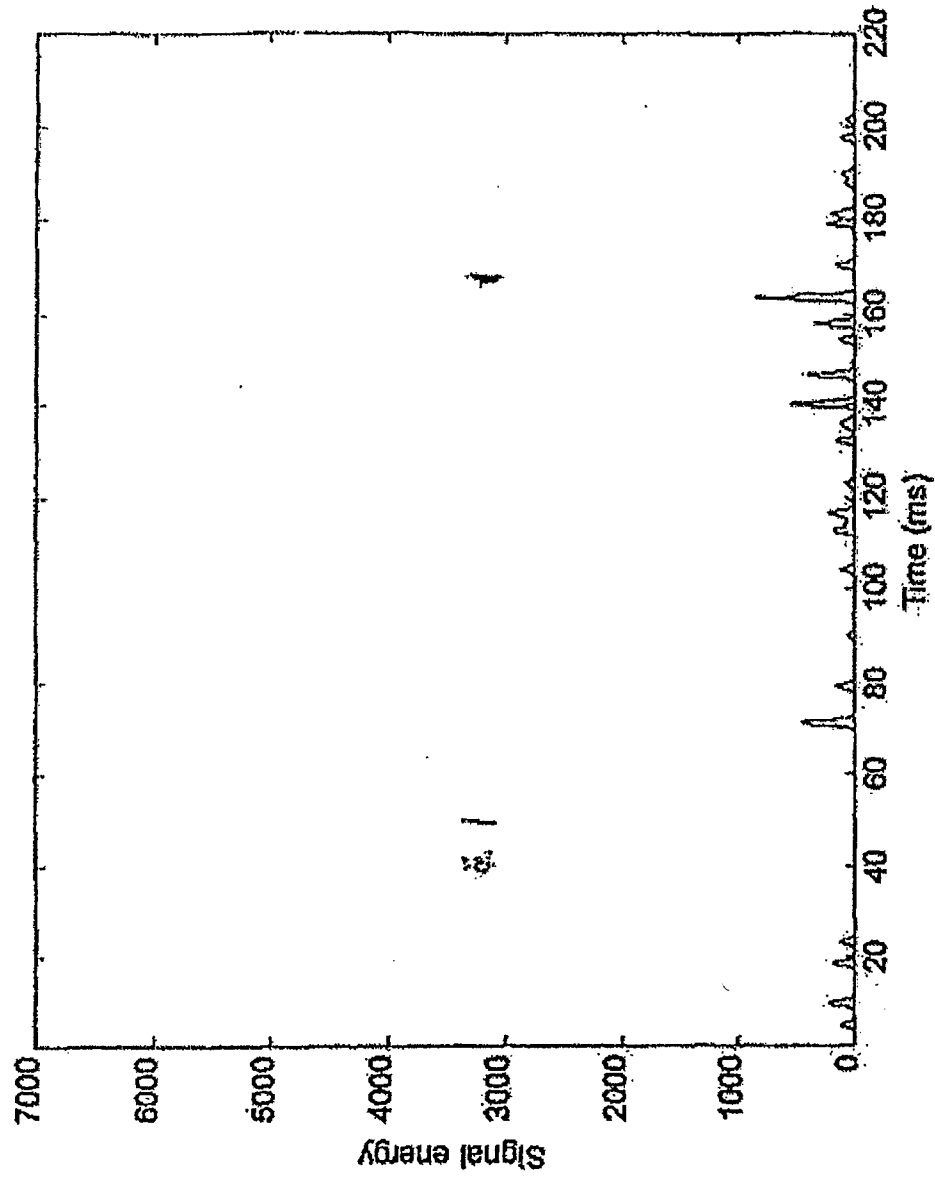


FIG. 12